

2015

A Case Study Examining Formative Assessment in a Postsecondary English Language Program

Divya Varier

Virginia Commonwealth University, varierd@vcu.edu

Follow this and additional works at: <http://scholarscompass.vcu.edu/etd>

 Part of the [Bilingual, Multilingual, and Multicultural Education Commons](#), [Educational Assessment, Evaluation, and Research Commons](#), [Educational Psychology Commons](#), and the [Higher Education Commons](#)

© The Author

Downloaded from

<http://scholarscompass.vcu.edu/etd/4016>

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

© Divya Varier 2014

All Rights Reserved

A Case Study Examining Formative Assessment in a Postsecondary English Language
Program

A dissertation submitted in partial fulfillment of the requirements for the degree Doctor of
Philosophy at Virginia Commonwealth University.

by

Divya Varier
B.A., Bangalore University
B.S., Old Dominion University

Director: Dr. James H. McMillan, PhD
Professor, Department of Foundations of Education

Virginia Commonwealth University
Richmond, Virginia
December, 2015

Acknowledgement

I would like to thank many people who have been instrumental in the completion of this dissertation.

First, I thank my dissertation chair, Dr. James H. McMillan, for his guidance and support throughout my doctoral journey. I thank Dr. Lisa Abrams for her mentorship and for the many learning opportunities over the last four years that have shaped my research and evaluation knowledge. I thank Dr. Zachary Goodell and Dr. Sharon Zumbrunn for their guidance as committee members. I am grateful to the teachers and students who participated in the study for their commitment. I thank my peers in the doctoral program at the VCU School of Education for their support. In particular, I thank Ms. Erika Dumke for her role in the study and for being a source of support throughout. I also thank Dr. Shelby Forbes of Rescue Edit LLC for editing and proofreading the first four chapters of the dissertation.

I express my gratitude to my family without which I could not have reached this milestone: my father, K.V.R. Varier, and my sister, Vidya Raghu, for their unconditional love and support; my mother-in-law, Prabha Lokesh, for taking leave from work to care for me and my baby during those arduous months of new motherhood. I wish I could share this achievement with my mother, Ramani, who would have been proud of me. I lost her nearly twenty years ago, but her wit, wisdom, kindness, and love has brought me this far and will continue to influence every aspect of my life.

Most of all, I am indebted to my husband, Paurav Lokesh, for his willingness to support my academic pursuits. I thank him for being my source of strength, for his unceasing confidence in my ability, and for his patience through the ups and downs of this doctoral journey. I dedicate this dissertation to him.

Table of Contents

Table of Contents	iii
List of Tables	v
List of Figures	vi
Abstract	vii
Chapter 1: Introduction	1
Background for the Study	1
Overview of the Literature.....	2
Theories of Formative Assessment.....	2
The Study of Formative Assessment in Postsecondary ESL Contexts.....	3
The Role of Metacognition in Formative Assessment.....	4
Rationale for the Study	6
Research Questions	6
Design and Methods	7
Definition of Terms	8
Chapter 2: Review of the Literature.....	10
Method for Review of the Literature	10
The History, Development, and Theories of Formative Assessment	11
Characteristics of Formative Assessment	11
Issues in the Study of Formative Assessment.....	15
Theories of Formative Assessment.....	20
Feedback Model for Enhancing Learning.....	24
Sociocultural Theory of Learning and Development.....	26
The Study of Formative Assessment Using Sociocultural Theory	32
The Study of Formative Assessment in Higher Education ESL/EFL Contexts	35
Studies on Feedback Interventions	36
Teachers' Feedback Practices	37
Student Perceptions of Formative Feedback.....	43
Formative Assessment Practices in ESL Contexts	45
Synthesis	51
The Role of Metacognition in Formative Assessment.....	52
The Relationship between Metacognition and Self-regulated Learning.....	55
A Review of the Theory and Applied Research of Metacognition.....	59
Application of Metacognitive Judgments of Learning	61
Synthesis	70
Chapter 3: Methodology	73
Research Questions.....	74
Study Overview	74
Sampling Procedures	74
Participants.....	77
Context.....	81
Assessments	83

Measures	83
Procedures	90
Data Analysis Procedures	92
Validation Procedures	95
Institutional Review Board	97
Ethical Considerations	98
Study Delimitations	99
Chapter 4: Results	100
Quantitative Results	101
Qualitative Results	112
Coding and Analysis Procedures	112
Categories and Themes	118
Contextual Characteristics	120
Teacher Characteristics	124
Student Characteristics	129
Formative Assessment	137
Synthesis	157
What are the ongoing formative assessment practices of teachers in ESL classrooms?	158
How did the three teachers use classroom assessment to inform instruction?	165
How did the three ESL teachers use the metacognitive data they received in their formative assessment practice?	167
Chapter 5: Discussion	172
The Relationship Between Judgments of Learning and Performance	173
No Change in Students' Metacognitive Beliefs Related to Writing	178
The Nature of Teachers' Engagement in Formative Assessment	180
Limitations	187
Study Design and Implementation Related Issues	187
Analysis and Interpretation of Findings	190
Implications of Findings for Research	191
Conclusion	195
References	197
Appendices	213
ESL Teacher Assessment Practices Interview Protocol	214
ESL Teacher Assessment Practices Interview Protocol – II	218
ESL Student Metacognitive Beliefs Survey	221
ESL Formative Assessment Observation Guide	224
Metacognitive Judgments of Learning Questionnaires	226
JOL-Performance Summary Report	227
List of Codes	234
VITA	241

List of Tables

	Page
Table 1: Variations of Formative Assessment Characteristics	14
Table 2: Four Levels of Feedback described by Hattie & Timperley	25
Table 3: Description of Teacher Participants	79
Table 4: Description of Student Participants	81
Table 5: Sample Size for student surveys and JOLQ	102
Table 6: Bivariate Correlations Between Judgments of Learning and Unit Test Scores for Thomas and Daniel's Students	105
Table 7: Correlations between JOL and Test Scores for Linda's Students	108
Table 8: Descriptive Statistics and Paired Sample <i>t</i> -test Results of Pre and Post Student Survey	121
Table 9: Description of Categories and Themes for the Coding and Analysis of Interviews and Observations	118
Table 10: Feedback Practices of Teacher Participants Coded from Observed Teacher-Student Interactions	152
Table 11: Examples of Feedback Related Interactions for the Three Teacher Participants	153

List of Figures

	Page
1. Model of Formative Assessment Theory and Research	71
2. Model of Formative Assessment Theory in this study	72
3. Mixed Methods Design Procedural Diagram	76
4. Summary of the Coding and Analysis Process	113
5. Illustration of Observed Feedback Related Interactions	151
6. Formative Assessment Practice of Teacher Participant Thomas	163
7. Formative Assessment Practice of Teacher Participant Daniel	164
8. Formative Assessment Practice of Teacher Participant Linda	165

Abstract

A CASE STUDY EXAMINING FORMATIVE ASSESSMENT IN A POSTSECONDARY ENGLISH LANGUAGE PROGRAM

By Divya Varier, B.A., B.S.

A dissertation submitted in partial fulfillment of the requirements for the degree Doctor of Philosophy in Education at Virginia Commonwealth University.

Virginia Commonwealth University, 2014

Major Director: Dr. James McMillan, Professor, Foundations of Education

This study examined the formative assessment practices of three teachers in English as a Second Language (ESL) classrooms using a sociocultural theoretical framework. The study was conducted in a postsecondary ESL setting at a large public university in the southeastern United States. Using an embedded mixed methods design, this study employed teacher interviews and classroom observations to address the overarching question: What individual and contextual factors are present in the formative assessment practices of participant ESL teachers? The study also explored the relationship between student metacognitive judgments of learning (JOL) and performance with the purpose of informing formative assessment practice. To this end, 51 students responded to pre and post surveys on their metacognitive beliefs and judgments of learning questionnaires prior to three unit tests. Summary reports of students' JOL were provided to teachers for their review and use. Findings showed teachers in this ESL setting engaged in a variety of formative assessment techniques; successful implementation of their techniques were influenced by their instructional style and student attributes like attendance,

class participation, and students' academic or educational experiences. Findings also indicated the central role of assessments in this context that provided ample opportunity for formative assessment. Overall, findings point to the value of using a sociocultural theoretical lens to examine the nature of factors affecting teachers' formative assessment practice. With regard to the use of metacognitive judgments of learning in formative assessment, findings showed a mixed relationship between student JOL and performance, and there was no change in students' metacognitive beliefs about writing over the duration of the semester. Although teachers did not use the JOL information in their instruction, they attributed inaccuracies in judgments to students' achievement level. These findings are limited by implementation issues and sample size. Further study is needed to understand the nature of postsecondary ESL students' JOL in authentic assessment situations and their applicability in the formative assessment process.

Chapter 1: Introduction

Background for the Study

From preschool to graduate school, students are subject to a variety of assessments that inform teachers, parents, admission committees, and governments on their knowledge and skills. Assessments are indispensable for educators to make decisions about the placement, progress, and promotion of students. Standards-based reform has made testing and assessment a major component of K-12 education, and student success in postsecondary education is predominantly determined by test grades. Although there are several classifications of assessments, two main types are discussed in the context of education: summative and formative assessment.

Summative assessments are used to measure learning. They are typically administered at the end of an academic semester or year to discern whether and to what extent students have mastered concepts taught during the course; thus, summative assessments are primarily evaluative in their purpose. This type of assessment is also considered a high-stakes assessment because key decisions such as course grades, admissions, promotions, and evaluations of teacher and school performance are based on how students fare on these tests. Examples of summative assessments include standardized annual examinations in schools and mid-term and final examinations in college courses. Formative assessments, on the other hand, refer to a process in which assessments are used to inform instruction and support student learning. They are used by teachers to make instructional decisions, and by students, to become aware of their progress. Formative assessments can be formal in that they serve the diagnostic purpose of identifying gaps in understanding, or informal as they are embedded in classroom activities to elicit student understanding.

Both formative and summative assessment are germane to the area of educational assessment and measurement, an area which is concerned with the process of identifying and developing assessments and measurements of attributes, ability, and knowledge in educational contexts. Traditional approaches to assessment focus on gaining technical accuracy, so that test scores represent valid and reliable measures of skill, attitude, or ability.

In addition to technical accuracy, contemporary understandings of assessment focus on proving the validity of an assessment or the appropriateness of its intended use. For example, an arithmetic test for sixth graders must only include items that reflect sixth grade curriculum and instruction, and students' test scores may only be used as a measure of learning in sixth grade arithmetic, not as indicators of overall intelligence or innate mathematical ability. Assessment administrators must consider the consequences that result from using test scores for a specific purpose. For example, if the arithmetic test results are used to place students in remedial, regular, or advanced mathematics classrooms, are the consequences of grouping based on this test not only appropriate but beneficial for students? Where test use is clearly articulated, well-developed assessments have immense benefits for users, as seen in many standardized summative tests. With formative assessments, however, this measurement-based approach has proven limiting and insufficient.

Overview of the Literature

Theories of Formative Assessment

Formative assessment developed from traditional views of educational assessment that focused on the development and administration of periodic formal tests, which were labeled "formative." In the study of formative assessment, researchers neglected the process elements of teachers using assessment information, modifying instruction, and noting changes in students'

learning. Shepard (2000) argued that “if instructional goals include developing students' metacognitive abilities, fostering important dispositions, and socializing students into the discourse and practices of academic disciplines, then it is essential that classroom routines and corresponding assessments reflect these goals as well” (p. 8). Learning is a cultural-historical activity, where students' learning takes place in the context of cultural expectations (Elwood, 2007). The relationship between a student's culture and expectations and a school's is important for student success, and assessments should be interpreted bearing in mind the context formed by the backgrounds and experiences of students and teachers who are part of it. Any study of formative assessment practices or interventions must take these sociocultural influences into account. Such a view holds important implications for designing studies that include multiple methods (qualitative and quantitative) and multiple data sources (students, teachers, curriculum- and policy-related documents, etc.), and are aimed at providing a context-based interpretation of even specific formative assessment practices (Schoen, 2011).

The Study of Formative Assessment in Postsecondary ESL Contexts

The majority of formative assessment research is based on K-12 settings. Formative assessment has been in the foreground of K-12 educational assessment policy in international contexts like the United Kingdom, Singapore, and Australia. Compared to the vast research base in K-12 school settings, not much is known about the practice of formative assessment in U.S. college populations (Yorke, 2003), specifically in English as a Second Language (ESL) learning contexts. There are a few studies that have examined teachers' formative and classroom assessment practices using a sociocultural framework, but these studies have been conducted outside of the United States (Birjandi & Tamjid, 2012; Chen, May Kleowski, & Kettle, 2013a, 2013b; Cheng, Rogers, & Wang, 2007). Although postsecondary English language classrooms

have been discussed in the context of certain aspects of formative assessment research (Leki, Cumming, & Silva, 2008), such as feedback and self-assessment, a substantial portion of the discussion on postsecondary English language classrooms has been conducted within the narrow literature of second language acquisition (Leki et al, 2008; Panova & Lyster, 2002). In examining the literature on writing feedback interventions in second language college writing courses, studies on teachers' writing feedback practices, and students' perceptions of writing feedback in postsecondary ESL classrooms, one can note the need to better understand the contextual, social, and individual factors at play in ESL/EFL writing. Formative assessment provides a sound theoretical framework to facilitate such understanding.

The Role of Metacognition in Formative Assessment

The literature on formative assessment has predominantly focused on the abilities, skills, and knowledge that teachers use to engage in formative assessment. Little is known about how students' cognitive and motivational characteristics impact the practice of formative assessment. Even the best feedback or instructional strategy does not facilitate learning if students do not act on the feedback or instruction they receive (Sadler, 1989). What this means is that the cognitive and motivational characteristics of students influence the learning process. The contribution of student metacognition, motivation, and perception toward learning and assessment in formative assessment practice has yet to have been examined. Metacognition is a cognitive component of student understanding that refers to an individual's ability to monitor and control cognitive processes like memory and learning strategies (McCormick, 2003). Successful learning includes metacognitive processes such as being aware of learning difficulties, evaluating whether learning has occurred, and knowing and using appropriate strategies to bridge the gap between current understanding and learning goals. A vast amount of empirical work has indicated that

metacognitive awareness is an important factor in student learning and achievement (McCormick, 2003), and metacognitive skills can be taught through instruction and practice (Winne & Perry, 2000). Additionally, interventions are often designed to provide explicit instruction in metacognition, which benefits a broad range of students (Veenman, Hout-Wolters, & Afflerbach, 2006). Most of the studies on metacognition, however, have been conducted as special interventions or experimental studies, outside authentic classroom settings. There is a need to incorporate what we know about metacognition and its role in student learning into instruction, and formative assessment provides an avenue for doing just that.

Andrade (2010) identified recent developments in self-regulated learning research as contexts in which we should consider classroom assessment. Metacognitive knowledge is an important aspect of self-regulated learning that has implications for formative assessment. Monitoring one's performance and incorporating external feedback are essential to successful formative assessment. For example, student reports on their thought processes during assessment tasks can provide useful information to teachers on their current understanding. Teachers can gauge whether student reports on their preparation, knowledge, perception of difficulty, and confidence regarding assessment content are consistent with their performance. Inconsistencies may be indicative of misconceptions at the individual and/or class level.

Metacognitive knowledge has been studied extensively using student self-reports on learning and student predictions about their performance on assessments. These measures are known as judgments of learning and typically include reports of student confidence level, perceived difficulty of the learning material or assessment, and predictions of performance on assessments. A majority of studies have documented patterns in students' judgments of learning, primarily in laboratory or experimental settings (Hattie, 2013; McCormick, 2003). Researchers

have also applied metacognitive monitoring measures to classroom contexts (e.g., Huff & Nietfeld, 2009; Miller & Geraci, 2010). In general, students tend to be overconfident in their judgments of learning in relation to their performance and demonstrate stability in the accuracy of their judgments over time. High-performers are consistently more accurate in their metacognitive accuracy, whereas low-performers consistently make inaccurate judgments. Further, there are preliminary indications that students' metacognitive knowledge and judgments may be related to feedback-seeking behaviors, and that feedback on students' accuracy judgments may help them improve their accuracy. However, there is no research examining how metacognitive accuracy can be used by teachers in their feedback practices to help students improve their learning.

Rationale for the Study

Designed for college-level ESL writing teachers and learners, this study represents a first step in addressing three gaps in the current research: 1) the gap in the use of sociocultural theory to conceptualize formative assessment studies; 2) the gap in studies on formative feedback in ESL writing, which have largely ignored the interaction between contextual and individual factors; 3) the gap in the potential application of well-established metacognition measures as a means to gather information on student thinking in the context of formative assessment.

Research Questions

Quantitative Questions:

RQ1. What is the relationship between student metacognitive judgments and student performance on unit tests?

RQ2. What is the difference between student metacognitive beliefs at the beginning and the end of the semester?

Qualitative Questions:

RQ3. What ongoing formative assessment practices do teachers use in the ESL classroom?

RQ4. How do teachers in the ESL classroom use classroom assessment to inform instruction?

RQ5. How do ESL teachers use data on metacognitive judgments of learning in their formative assessment practices?

Design and Methods

Using a sociocultural theoretical framework entails an examination of the learning context, its participants, and the interactions among them through in-depth case studies. The identity and agency of both learner and teacher are critical to the formative assessment process. Qualitative information on participant experiences obtained by interviews, and documenting interactions through observations, captures perceptions and processes related to formative assessment in this specific context. This study uses an embedded mixed methods design (Creswell & Plano Clark, 2011) that incorporates qualitative interview data from teachers, quantitative survey data from students, and performance data including unit test scores and the final course grade.

Data collection methods included two student questionnaires (pre and post survey on metacognitive beliefs, influences on writing, and application of coursework in writing; one judgments of learning questionnaire administered prior to each unit test), two interviews (primary and follow-up), and eight non-participant classroom observations (two observations for each class/section).

Prior to implementation, this study was submitted to the Virginia Commonwealth University Institutional Review Board (Approval Number: HM20002382) for approval for an expedited review. The current study was conducted for the duration of one academic semester (fall 2014) in four ESL classrooms focusing on written communication. The students featured in this study were classified as level 3 advanced English learners. Upon completion of this course, students are eligible to enroll in undergraduate academic coursework at a U.S. university. A majority of students in the program are enrolled full-time. The English language program follows a common curriculum by requiring all classes to use *Q- Skills for Success* learning materials, conduct four unit tests (end of unit), as well as administer one mid-term and one final examination. Although most assessments are somewhat standardized across classes, instructors may exercise flexibility in modifying specific tasks for unit tests.

Definition of Terms

1) *Formative assessment*: Formative assessment is the process by which teachers elicit and make sense of evidence of student learning, both to inform instruction and support learning in order to achieve learning goals. Formative assessment includes ongoing informal assessments in daily instructional interactions, as well as formal assessments administered periodically throughout the course of learning.

2) *Classroom assessment*: Classroom assessments are formal teacher-made assessments used to provide information on student progress; this type of assessment is not a high-stakes assessment in that it does not greatly affect learners' individual grades for the course. Classroom assessments are administered at the end of one or more instructional unit/s.

3) *Sociocultural theory of learning and development*: This theory is based on a philosophy that considers individual, social, and contextual influences on human activity (Schoen, 2011). It is

rooted in the notion that human behavior, motivation, and learning are shaped by the complex interactions between the individual and the social world.

4) *Metacognition*: Metacognition refers to learners' cognitive awareness of the skills they have in place for learning, and their capability to activate, use, and modify such skills (McCormick, 2003).

5) *Metacognitive judgments of learning*: This term refers to students' perceptions of their degree of preparedness, their confidence in an academic skill or task, and the difficulty level of learning material. The measurement of metacognition through self-report typically involves one or more of the following: confidence judgments, learning judgments, difficulty/ease judgments.

Confidence judgments refer to the extent to which the learner is confident that his or her response is correct.

6) *Self-regulated learning*: Self-regulated learning is a cyclical process that involves planning or forethought before taking on a task, monitoring and making adjustments during learning, and lastly, evaluating the resulting performance, known as the self-evaluation phase (Zimmerman & Cleary, 2004).

7) *English as a second language (ESL)*: English as a second language refers to English language classrooms where the primary beneficiaries of instruction are speakers of other languages, who learn English as an additional or second language. It is distinguished from *English as a foreign language* (EFL), a term which refers to English language classrooms located in countries where English is not the native language.

Chapter 2: Review of the Literature

Method for Review of the Literature

The scholarly literature included in this review was identified primarily through online searches using Google Scholar and VCU Libraries. Search terms included the following descriptors, used separately and in combination, without restricting publication dates: *formative assessment, formative assessment theory, higher education/college, EFL/ESL, feedback, writing feedback, metacognition, metacognitive judgments, and sociocultural theory*. A typical two-descriptor combination yielded between 43 (*formative assessment theory + higher education OR college*) and 6,070 results (*ESL OR EFL + formative assessment*). An even narrower search, such as *formative assessment + metacognition + college OR higher education*, yielded 7,440 results. In order to ease the selection process, the following strategies were employed: 1) the first three pages were skimmed after sorting results by relevance; 2) article titles that did not directly relate to the search (e.g., work environments, special populations) were excluded; 3) articles with more citations were often shortlisted; 4) in Google Scholar, articles referenced by widely cited articles were searched (up to 5 pages). The following databases provided access to the majority of the journal articles in this review: ERIC, APA PsycNET, EBSCO, Taylor & Francis Journals, JSTOR, Springer Link, and, occasionally, Science Direct. The following journals, known to publish empirical and theoretical research on formative assessment, metacognition, sociocultural theory, and higher education, were searched individually for relevant articles: *Educational Assessment: Principles, Policy and Practice; Higher Education; Studies of Higher Education, Assessment & Evaluation of Higher Education; TESOL Quarterly; Language Assessment Quarterly; and Metacognition & Learning*. A specific search including all the main descriptors (*metacognition + formative assessment + EFL + sociocultural theory*)

yielded 61 results on Google Scholar. Relevant studies were selected for the review. Scholarly books were identified based on the following: 1) recommendations by committee members; 2) citations in journal articles; and, 3) database search results.

This review of the literature is comprised of three sections: 1) Theories of formative assessment, which includes the history and development of formative assessment theory, characteristics of formative assessment, and issues related to the empirical research on formative assessment leading up to sociocultural theory. This section identifies a theoretical gap in the research on formative assessment, which supports the use of a sociocultural framework to examine formative assessment; 2) Formative assessment in higher education, which describes empirical work on formative assessment conducted in postsecondary classrooms, specifically, English as Second Language (ESL) contexts, and identifies the empirical gap in formative assessment practices; 3) The role of metacognition, which describes current understandings of metacognition, particularly metacognitive learning judgments that provide insight into learners' cognitive processes during learning; this section also describes studies on formative assessment in educational contexts that examined the role of student metacognition, perceptions, and beliefs in relation to their use in instructional practices.

The History, Development, and Theories of Formative Assessment

Characteristics of Formative Assessment

The term “formative assessment” can be traced back to the use of the terms “formative” and “summative evaluation” by Scriven around 1967 in the context of program evaluation. Formative assessment was popularized when the term was used by Bloom (1971) in his *Handbook of Formative and Summative Evaluation of Student Learning* (as cited in Cizek, 2010; Guskey, 2005). Bloom makes a distinction between formative and summative by referring to

summative evaluation as assessments that occur at the end of an instructional unit, and formative evaluation as assessments that occur during instruction, or in the process of learning. As in program evaluation, the purpose of summative assessment is to aid decision making in order to provide a judgment as to whether or not learning has occurred. The purpose of formative evaluation was to improve learning, and a critical component of formative evaluation involved the removal of the evaluative component. Bloom conceptualized the solution to reducing the variability in student learning or achievement as lying with the introduction of variability in instructional strategies, indicative of formative assessment principles. Since Bloom, formative assessment has become less associated with program evaluation and an integral part of educational assessment (Torrance, 1993).

The three steps that form the foundation of formative assessment have remained the same over the last 30 years: knowing (1) where the learner needs to be, (2) where the learner is, and (3) what needs to be done to get him or her there (Brookhart, 2007; Hattie & Timperley, 2007; Ramaprasad, 1983; Sadler, 1989; Wiliam, 2010). The first step entails a clear understanding of learning goals, and the identification of criteria that demonstrate the achievement of learning goals. In the second step, evidence of current understanding is elicited through instructional activity or assessment. In the final step, the elicited information is then used to determine whether learning has occurred, what needs to be done, and which strategies and correctives are required to close the learning gap. This three-step process may be spontaneous or planned, formal or informal, include oral or written communication, conducted at the individual- or group-level, and supported by a teacher, peer, or computer. In any classroom setting, teachers may employ, deliberately or otherwise, a combination of formative assessment strategies. Therefore, formative assessment characteristics in practice have been described in several ways.

Black and Leahy (2007) characterized five strategies of formative assessment based on Ramaprasad's (1983) feedback framework: 1) sharing learning objectives with learners; 2) eliciting evidence of learning through discourse; 3) providing feedback; 4) activating peer assessment; and 5) activating self-assessment. Formative assessment, according to this framework, is an ongoing process that includes daily instructional practices in the classroom. Similarly, McManus (2008) defined formative assessment as a process where evidence of learning "is used by teachers and students to inform instruction and learning during the teaching/learning process" (p. 3). McManus provided a list of five attributes of effective formative assessment: (1) Instruction should be firmly based on learning progressions or learning trajectories as established by content domains. This enables teachers and students to work on clear gaps to inform feedback and instructional correctives. (2) Learning goals and criteria for successful learning are articulated and communicated to students since it is important for students to know the end objective as they regulate their learning. (3) Specific, timely, and descriptive feedback on student performance should be given to enable students to identify learning gaps as well as know how to close the gap. (4) Instruction should utilize peer- and self-assessment as tools to encourage students to think about their learning or metacognition. Providing and receiving feedback can support student reflection and improve his or her understanding of the criteria used to evaluate the work and the quality of work expected. Similarly, self-assessment promotes students' perceptions of their learning and fosters self-regulation. (5) Students and teachers act as collaborative partners in learning.

Formative assessment has also been interpreted along a continuum based on the extent to which these three steps are implemented in an assessment activity (McMillan, 2010). Table 1 provides a description of teacher, student, and contextual characteristics that represent low,

medium, and high levels of formative assessment. Table 1 has been reprinted from McMillan (2010), and it describes the variations in formative assessment characteristics interpreted as low, moderate, and high level formative.

Table 1

Variations of Formative Assessment Characteristics

Characteristics	Low level Formative	Moderate-level Formative	High-level Formative
Evidence of student learning	Mostly objective, standardized	Some standardized and some anecdotal	Varied assessment, including objective, constructed response, and anecdotal
Structure	Mostly formal, planned, anticipated	Informal, spontaneous, “at the moment”	Both formal and informal
Participants involved	Teachers	Students	Teachers and students
Feedback	Mostly delayed and general	Some delayed and some immediate and specific	Immediate and specific for low achieving students, delayed for high achieving students
When done	Mostly after instruction and assessment	Some after and during instruction	Mostly during instruction
Instructional adjustments	Mostly prescriptive planned	Some prescriptive, some flexible, unplanned	Mostly flexible, unplanned
Choice of tasks	Mostly teacher determined	Some student determined	Teacher and student determined
Teacher-student interaction	Most interactions based primarily on formal roles	Some interactions based on formal roles	Extensive, informal, trusting, and honest interactions
Role of student self-evaluation	None	Tangential	Integral
Motivation	Extrinsic	Both extrinsic and intrinsic	Mostly intrinsic
Attributions for success	External factors	Internal stable factors	Internal unstable factors

Note. Variations of Formative Assessment Characteristics. Reprinted from “The practical implications of educational aims and contexts for formative assessment” (p. 43), by J.H. McMillan, 2010, *Handbook of Formative Assessment*, by H. Andrade and G. Cizek (Eds.), New York: Routledge.

Formative assessment has also been described in relation to evidence from formal assessments used to inform instruction (Wiliam & Thompson, 2007 as cited in Brookhart, 2007). Short-cycle formative assessments occur within lessons on a daily basis in the classroom. Medium-cycle formative assessments occur between instructional units, and are designed to provide evidence of student understanding and inform instructional decisions. Long-cycle assessments occur across multiple units, or at the end of a semester or year (Brookhart, 2007; Supovitz, 2012). The question of how the use of this assessment improves learning is critical to this interpretation of formative assessments.

More recently, Brookhart (2013) provided a framework to understand the function of assessments along the dimensions of administration- (classroom-based to large scale), and the purpose of assessment(formative to summative). The closer assessments are to classroom-based administration with a formative purpose, the more likely they are to be used to support learning and make instructional adjustments.

The broad range of interpretations of formative assessment has led to a diverse body of literature on the subject. Diverse perspectives have called into question the extent to which such vast empirical research collectively contributes to our understanding of formative assessment (Bennett, 2011; Dunn & Mulvenon, 2010; Good, 2011). The following section elaborates on said reviews of formative assessment literature.

Issues in the Study of Formative Assessment

Bennett (2011) discussed issues in formative assessment research, particularly, definition-related issues pertaining to whether formative assessment is an instrument or process. Testing companies provide item banks and practice tests labeled “formative assessment” with the intention of providing teachers with short- or medium-cycle assessments for tracking student

progress with respect to standardized assessments. Theoretical conceptualization of formative assessment (Sadler, 1989; Shepard, 2000) has focused on the process of a feedback loop, student self-assessment, and dialogic interactions between teachers and students. Another body of literature (McManus, 2008; Supovitz, 2012), however, conceptualizes formative assessment as a process where instructional adjustments are made based on student performance where any instrument, including summative assessments, can serve a formative purpose. This has led some researchers to replace the use of the phrase “formative assessments” with “assessments for learning,” and “summative assessments” with “assessments of learning,” to clarify the distinction between the supportive and evaluative function of assessments. Bennett argued, though, that replacing formative assessment with assessment for learning does not remedy the oversimplification of the concept. Bennett also reviewed widely cited studies on the effectiveness of formative assessment in learning. He found that a majority of studies, especially research syntheses and meta-analyses, analyzed disparate studies that did not necessarily reflect formative assessment as conceptualized for classroom contexts.

In the most widely known research synthesis on formative assessment, Black and Wiliam (1998) examined 681 empirical studies of formative assessment, and concluded that formative assessment showed larger gains in terms of student performance than any other educational intervention. They noted that the effect size of student performance improvement could be anywhere between 0.4 and 0.7, which is considered substantial for educational interventions. The authors reviewed articles from a diverse body of education research that could be identified as formative assessment. Black and Wiliam’s review is arguably the most cited in the formative assessment literature. However, researchers recently examined the empirical studies in the review, and found that they were too disparate and methodologically unsound to draw

meaningful conclusions on the efficacy of formative assessment (Bennett, 2011; Dunn & Mulvenon, 2009).

Early research on feedback, an important component of formative assessment, was rooted in behaviorist theories of learning, where feedback was considered a form of reinforcement. Studies examining the effects of feedback were conducted using artificial learning tasks in laboratory-like settings (Wiliam, 2013; Ruiz-Primo, 2013). Several meta-analytic studies examining the effects of feedback have been cited in support of formative assessment. The importance of the nature and content of feedback have been found to have differential effects on performance. For example, a widely cited study by Kluger and DeNisi (1996) included 607 effect sizes of feedback interventions on performance, and found an average positive effect of 0.4 standard deviations. The study provided important insight into which feedback interventions are supportive to performance and which are detrimental; however, the interventions did not represent contexts of teaching and learning. In addition, performance outcomes included several non-academic tasks that call into question the validity of conclusions made regarding the effectiveness of formative assessment. In spite of these limitations, Kluger and DeNisi identified important characteristics of feedback, and its relationship to the learner's internal metacognition, motivation, and personality, which can, in turn, provide valuable information to the theory of formative assessment.

Contextualizing feedback with respect to classroom assessment, Ruiz-Primo (2013) noted that there is a dearth of research examining the nature of feedback in ordinary classroom settings and within regular classroom interactions. Feedback, like learning, is a social process, and occurs within teacher-student interactions. In terms of theoretical contribution, it is more valuable to examine feedback practices in different classroom contexts, while taking into account

how student characteristics may influence feedback practices and the efficacy of feedback in different learning contexts.

A more recent quantitative meta-analysis conducted by Kingston and Nash (2011) refuted Black and Wiliam's (1998) findings, asserting that there was an overestimation of the effect of formative assessment, and, instead, reported a smaller effect size of 0.3. However, even this rigorous meta-analysis was wrought with issues similar to those characterizing Black and William's analysis (Briggs et al., 2012; McMillan, Venable & Varier, 2013). The empirical articles Kingston and Nash consulted were all quantitative, experimental, or quasi-experimental designs, yet they investigated different aspects of formative assessment, from professional development to computer-based feedback, thus calling into question both the quality of empirical studies and the appropriateness of conducting meta-analyses. Another argument against the reported impact of formative assessment cited by meta-analyses on formative assessment concerns differences in the implementation of formative assessment in empirical studies. Many studies of formative assessment do not adequately measure whether, to what extent, and in what ways teachers implemented the assessment process in the classroom (e.g., Wiliam, Lee, Harrison, & Black, 2004). In other instances, formative assessment has been studied without sufficient consideration of contextual factors such as the policy surrounding learning and assessment or pressures to perform well on standardized large scale assessments (Abrams, 2007; Shepard, 2000).

The majority of formative assessment research has been conducted in K-12 settings (Yorke, 2003). Several issues in formative assessment research, including its definition, interpretation, and study, may be attributed to the learning and assessment context of K-12 education. Torrance (1993) claimed that in the UK, formative assessment was considered an

effective tool for combating the negative effects of large-scale assessments, despite the lack of empirical or theoretical evidence speaking to its efficacy within the context of accountability-based testing. He noted that formative assessment was implemented as a way to obtain information on student progress at the national level curriculum. Torrance argued that the use of formative assessment was “mechanistic” and “behaviorist” (p. 335-336) in contexts where all teachers and students work toward achieving objectives demonstrated through performance in standardized assessments. By making assumptions about the effectiveness of formative assessment at the outset, a key question regarding its role in the learning process has been left unanswered.

In the United States, Abrams (2007) reviewed empirical literature on the impact of high-stakes testing on learning, achievement, and instruction. She reported several detrimental effects on teacher morale, as well as increased student anxiety, increased pressure to cover a breadth of content at the expense of depth, and increased instruction of test preparation skills to improve scores on state tests. High-stakes testing led formative assessment practice to be reinterpreted as any assessment that was administered between summative assessments, where the focus was no longer on the process of feedback and instructional correctives, but frequent testing. This interpretation was reinforced by the growth of commercial products aligned with state tests labeled and described as formative assessments.

The shift away from formative assessment as a process in learning and instruction also led to an interpretation of formative assessment as a unidirectional, teacher-led activity that impacts student learning, as demonstrated in summative assessments. Formative assessment research has been dominated by interventions like professional development programs for teachers (William et al., 2004), or feedback given to students (Boulet, Simard, & De Melo, 1990)

without regard to the role of students in the formative assessment process. Yet, the primary outcome of formative assessment studies has been student achievement on subject or standardized assessments. For example, all of the studies included in Kingston and Nash's (2011) meta-analysis were experimental or quasi-experimental studies with comparison groups. These studies also included various formative assessment interventions like computer-generated feedback, oral/written teacher feedback, or professional development for teachers. Although several of these studies found a positive effect on student achievement scores, they did not contribute to better understanding of the formative assessment process, i.e., how the intervention promoted changes in students' learning.

In summary, current issues related to formative assessment research may be attributed to several factors, including: diverse interpretations of what constitutes formative assessment; incorrect assumptions regarding its effectiveness, based on studies that do not reflect authentic learning contexts; a vast amount of empirical literature based on these assumptions; and the contextual influences of K-12 settings on the implementation and interpretation of formative assessment. In light of these concerns, the following section describes the development of major theories of formative assessment.

Theories of Formative Assessment

Ramaprasad's (1983) foundational article on feedback prompted him to develop a theory on the role of feedback in student learning. His ideas about understanding where the student ought to go in relation to where the student currently is, and methods for reducing the discrepancy between the two, was appealing to formative assessment research.

The most influential article on the theory of formative assessment was published by Sadler (1989), who drew upon Ramaprasad's work to provide a clear foundation for formative

assessment as a concept. According to Sadler, "Formative assessment is concerned with how judgments about the quality of student responses can be used to shape and improve students' competence by short-circuiting the randomness and inefficiency of trial-and-error learning" (p. 120). Sadler believed feedback, as well as student participation, to be critical to effective formative assessment. Feedback involves not only providing information about the learner's performance, but also making instructional adjustments to help the learner progress toward the goal/standard. For Sadler, learner participation involves student knowledge and ownership of a standard (i.e., acceptance of the standard), knowledge of his/her actual performance, and initiating action that would close the gap between actual and goal performance.

Sadler (1989) also described the role of self-monitoring, where feedback is not only external, but also internal, coming from within the student. In formative assessment, qualitative judgments are made about a student's performance, making the use of numbers/scores unnecessary, which is consistent with the original definition of formative assessment provided by Bloom (1971). The ultimate aim of formative assessment is to help students set and internalize learning goals through motivational, self-regulative, and metacognitive elements. Sadler's framework is widely accepted, and forms the basis for current conceptualizations of formative assessment. However, his emphasis on the role of students, and the importance of motivation and self-regulation, was undermined by subsequent empirical investigations and interpretations of formative assessment. Only recently, has theoretical work incorporated motivation theory and research in a way that places the learner at the center of the formative assessment process (Andrade, 2013; Brookhart, 2013).

Shepard (2000) viewed assessment, especially classroom assessment, as an integral part of instruction. In her influential article on the role of assessment in cultures of learning, Shepard

traced the evolution of pedagogy and assessment theory from a historical perspective. Theories about teaching and learning have evolved so as to consider the strong social and cultural influences on learning and development. Teaching and learning goals include providing authentic learning experiences, and fostering students' critical thinking and metacognitive skills. Shepard argued that assessments, however, have remained in the periphery of this change, and in form and function, assessments conform to traditional, behaviorist views of learning. She advocated the need to shift to a social constructivist view of assessment to reconcile this incompatibility between teaching goals and assessment. Shepard's article reflects an important realization about the vast discrepancy between the theorization of formative assessment as a contextualized, interactive process between teachers and students, and its empirical study and practice as a teacher-directed activity separate from instruction. The latter of these two views of formative assessment largely constitutes the body of literature on the subject matter. In addition to Shepard, other researchers have also noted that empirical research has failed to adequately inform theories of formative assessment (Kluger & De Nisi, 1996; Ruiz-Primo, 2013).

Black and Wiliam (2009) provided several opportunities for including learner cognition and motivation in theories of formative assessment. Their first theoretical framework defined formative and summative assessment by the function they served, in the case of formative assessment, whether or not it results in instructional adjustments that regulate learning for the individual (Black & Wiliam, 2009; Wiliam & Leahy, 2007). They extended this theoretical framework using several pedagogical and learning theories to incorporate the role of learners' internal processes critical to formative assessment. From here, they outlined several possibilities to account for learners' cognitive processes in order to address conceptual gaps in formative assessment theory.

Nicol and McFarlane-Dick (2006) combined a model of self-regulated learning, specifically, external feedback provided by teachers and peers, with the theory of formative assessment. They used the self-regulated learning model provided by Butler and Winne's (1995) which describes the fine-grained components and processes of self-regulated learning internal to the individual. An external cue in the form of tasks or goals activates the individual learner's prior knowledge, strategies, motivational and knowledge beliefs. The individual proceeds to plan and implement learning strategies, and monitors whether or not expectations for the task are being met, thus providing self-directed feedback. The resulting performance then leads to feedback from teachers or peers, which is classified as external feedback. Nicol and McFarlane-Dick posited that the interaction between internal and external processes determines the effectiveness of formative assessment. They extended the self-regulated learning model to incorporate connections between external feedback and internal processes, recommending seven principles of feedback that support self-regulated learning based on this model, including: (1) providing clear criteria in order to achieve good performance, (2) encouraging self-assessment, (3) delivering high-quality task-based feedback, (4) encouraging teacher and peer interactions, (5) fostering motivation and self-esteem, (6) providing opportunities for students to use feedback (revisions), and (7) using student performance data to improve and inform instruction.

Feedback, according to Hattie and Timperley (2007), refers to the "information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding" (p. 81). In a review of the conceptual elements of feedback and its impact on student learning, the authors noted several characteristics of teacher feedback, and constructed a model of feedback to enhance learning. Their view of feedback rejected behaviorist notions of feedback as a reinforcer for learning, and, instead, viewed feedback as an

active interaction between the teacher/peer and the learner. Per this model, the learner can seek feedback, and take action by accepting or rejecting said feedback.

Feedback Model for Enhancing Learning

In Hattie & Timperley's (2007) feedback model for enhancing learning, the purpose of feedback was to reduce the discrepancy between the learner's current understanding and the desired goal. The researchers posited that in a teaching and learning context, the discrepancy can be reduced by students themselves who may exert increased effort or use specific strategies, or by teachers who may use instructional strategies and feedback. Conversely, based on the feedback they receive, students may readjust their learning goals to reduce the gap between such goals and the feedback, or teachers may provide more appropriate goals. Effective feedback is comprised of four levels and addresses the three foundational principles of formative assessment: where the student is going, how the student will get there, and where the student will go next. Table 2 provides a description of each level of feedback. The first level, task-level feedback, is the most common form of feedback, and refers to specific feedback related to the correctness/incorrectness of a response. This type of feedback is the most basic, but if too specific, it may hinder learning by focusing students' attention on the immediate task and reducing cognitive effort. Task-specific feedback is most beneficial to highly strategic students with a greater ability to discern their learning. The second level includes process-level feedback, which refers to feedback on the processes underlying a learning task; this may include providing information on how to make corrections rather than providing correct answers. This level is more effective than task-based feedback in supporting learning since it encourages students to reflect on the concept or procedure to guide corrections or revisions.

Table 2

Four Levels of Feedback described by Hattie & Timperley

Levels of feedback	Description
Task level	How well tasks are performed
Process level	The main processes needed to understand/perform tasks
Self-regulation	Self-monitoring, directing, and regulating of actions
Self-level	Personal evaluations and affect (usually positive) about the learner

Note. From ‘a model of feedback’ illustration. Adapted from “Power of Feedback”, by J. Hattie and H. Timperley, 2007, *Review of Educational Research*, 77(1), p. 87. Copyright American Educational Research Association

The third level includes self-regulation feedback, which involves self-monitoring and self-appraisal of task performance. This internal feedback has important connections to learning. Students who have the metacognitive skills to self-assess are able to judge their learning in relation to others, as well as demonstrate accurate judgments of their effort, attributions, and learning. Hattie and Timperley (2007) also relate a student’s confidence in their own learning to his or her willingness to seek and accept feedback. The fourth level includes self-level feedback, an undesirable type of feedback. This level refers to feedback that provides information on the learner’s personal attributes, such as ability. By analyzing the vast amount of literature on feedback interventions, Hattie and Timperley concluded that self-level feedback, both positive and negative, is associated with a negative impact on learning.

Hattie and Timperley (2007) also noted that in classroom contexts, task- and self-level feedback were more common when compared to the more effective process-oriented feedback. In addition to these four levels, they addressed the effect of feedback based on timing, the nature of the feedback (positive or negative), and feedback provided in the classroom. As for feedback based on timing, they noted that for difficult tasks, delayed feedback was the most effective for supporting learning. Then, regarding the nature of the feedback, Hattie and Timperley found that negative feedback holds a much more powerful influence than positive feedback. For example,

students with high self-efficacy (beliefs about their ability to complete tasks successfully) may seek negative feedback to help them perform better at tasks, and use positive feedback to confirm self-efficacy beliefs.

The present study aimed to explore whether knowledge on how having access to students' judgments of their own learning can inform teachers' formative feedback and instructional practices. Along these lines, Andrade (2013) discussed the implicit role of metacognition in formative assessment. Although the papers on formative assessment discussed in this review make explicit reference to the importance of student beliefs, metacognition, and self-regulatory skills in the process as well as the outcome of formative assessment, the measures widely used in motivation and cognition research are "rarely taught much less assessed in classrooms" (p. 24). However, a number of measures exist to assess these student characteristics (Schraw, 2009; White & Frederiksen, 2005), measures which demonstrate the potential for informing feedback and instructional practices in the classroom.

The following section elaborates on sociocultural theory as a framework for investigating the role of student characteristics in teachers' formative feedback practices.

Sociocultural Theory of Learning and Development

Sociocultural theory has been in the foray of education research, particularly educational psychology, for the last 30 years, and is increasingly regarded as a viable theory of learning and, more recently, of motivation (McInerney, Walker, & Liam, 2011). Sociocultural theory is also known as cultural-historical activity theory (CHAT), activity theory, and sociocultural psychology, among others. In essence, *socioculturalism* can be defined as a philosophy that considers individual, social, and contextual influences on human activity (Schoen, 2011). It is rooted in the notion that human behavior, motivation, and learning are shaped by complex

interactions between the individual and the social world. Sociocultural theory is especially relevant to education because expectations for development are set by the culture/context, and knowledge is imparted and acquired as a result of social interactions between teachers and students (Edwards & Daniels, 2006). Socioculturalists argue that trying to understand learning or motivation solely from an individual or social perspective provides an incomplete picture, and that the power of sociocultural theory lies in its holistic approach to human behavior.

Researchers in education and psychology have “controlled” for the influence of socioeconomic status or teacher characteristics when examining an aspect of learning and development. Information processing and social cognitive theories account for the social and cultural contexts of development, but the fundamental belief is that the influence of context can be separated from child development. Piaget (1970), for example, saw development as stemming from within the individual as a result of his or her own discoveries and maturation. Bandura (1986), on the other hand, viewed learning and development as a reciprocal interaction between the external environment and the individual that is mediated by the individual’s behavior. Reflecting on the different attributions of learning and development, in his introduction to Vygotsky’s (1978) social-historical theory of cognitive development, Crain (2000) asks, “Can’t a theory assign major roles to both inner and outer forces?” (p. 213).

The development of sociocultural theory is attributed to the Russian psychologist, Lev Vygotsky (Crain, 2000). Unlike that which was proposed by Piaget and Bandura, Vygotsky envisioned a theory that explained all aspects of development. Although, because much of his work remained incomplete, and due to insufficient and inaccurate translations, the full scope of his theory has not yet been explored (Tudge & Scrimsher, 2010). Still, Vygotsky’s writings on learning and cognitive development in school settings have special relevance to education

research. His work opposed behaviorism, which studied human behavior as an independent action carried out by individuals in response to the environment. Vygotsky's theory views the role of the individual and external social influences as inseparable components in their contributions to cognitive development. Vygotsky explains psychological development as the result of knowledge and skills passed on by adults and competent others, through language and instruction, to developing individuals. Child development occurs in a historical and cultural context as a result of dialogic interactions with others. To communicate and interact with the world, humans use a variety of symbols like language, numbers, and maps.

[Vygotsky] stressed the importance of reconstructing the social, environmental, and cultural forms and conditions surrounding the behavior and consciousness, and its development. In so doing, he linked a natural science approach (i.e., studying behavior to assess mental functions) to a philosophical approach (i.e., reflection on social interactions and cultural forms of interaction) and used the merged method to study cognitive development. (Schoen, 2011, p. 15)

Mediation is a word often used to describe development through interaction with the world by way of physical and psychological tools (Poehner, 2011). Development does not occur through maturation alone, and an individual does not develop in isolation. What individuals learn and how they learn is determined by the historical and cultural context that surrounds them. Through mediation, learning occurs on a social plane before it is internalized. Consistent with other theories, the aim of sociocultural theories of learning is to foster independent learning and autonomy. However, there is a constant negotiation with the environment, making learning and development a dynamic process.

Learning takes place in the context of the expectations established by a culture. The relationship between a student's culture and expectations and a school's is important for student success. Furthermore, in interpreting development one bears in mind the context formed by the backgrounds and experiences of students and teachers who are part of it (Elwood, 2007). As an example of the importance of context to understanding development, Vygotsky (1978) made reference to secondary disabilities, which involved a social response to students with disabilities; such responses included low expectations for student performance and restricted access to quality schooling. Apart from the restriction imposed by an actual disability, the social reaction

to students with disabilities had far more negative consequences for students whose developmental trajectory was not a concern for adults around them.

Perhaps the most important term associated with Vygotsky's (1978) work is *zone of proximal development* (ZPD), which he defines as the difference/gap between independent performance and performance with the assistance of a competent peer or adult. In the context of formal learning, ZPD makes intuitive sense when considering teacher-student relationships. However, researchers in the Vygotskian tradition have criticized the reduction of sociocultural theory as a whole to just ZPD, which is commonly misconstrued in a quantitative sense, where the "zone" is a visible gap between independent performance and assisted performance of a skill.

A teacher's ultimate goal should be to make children autonomous learners. However, this goal is most successfully achieved when students and teachers have conversations about the learning process, and they both understand what is to be learned, how it is to be learned, and whether and to what extent learning has taken place (Scrimsher & Tudge, 2003). The role of the teacher is not that of an authoritative giver of knowledge, but a competent adult who is learning about children's conceptions and misconceptions, and engaging in dialogue to strengthen and expand their knowledge and skills. Teachers should help students bring what they know and see into "conscious awareness" (p. 298). The role of the assessor is not to remain at the periphery as "passive acceptors," but as "active modifiers" (Poehner, 2011, p. 102) who interact with the assessee to determine a developmental trajectory. From such a perspective then, the position of a teacher in formative assessment shifts from that of an evaluator to a collaborator who provides agency to the learner, and, in turn, learns from the learner.

Sociocultural theory provides a useful framework for understanding the theory and practice of formative assessment as a process in classroom contexts. The idea of mediated

learning has been well-established in learning interventions like dynamic assessment. Black and Wiliam (2009) proposed mediation as a useful concept for understanding how learning occurs in assessment-related interactions. Sadler's (1989) framework points to a similar notion of students acquiring the evaluative skills of an expert, who, in classroom contexts, is the teacher:

The indispensable conditions for improvement are that the *student* comes to hold a concept of quality similar to that held by the teacher...students have to be able to judge the quality of what they are producing and be able to regulate what they are doing during the doing it. (p. 121)

A key element of a sociocultural framework is the role of context/culture in shaping interactions and learning. Vygotsky explained that development does not occur within the individual removed from the context, but rather that the content and process of learning is determined by the culture in which the individual develops (Crain, 2000; Scrimsher & Tudge, 2003). The student's context may be shaped by teachers' assessment practices, which are influenced by their beliefs about teaching and learning, external factors in the form of policy and accountability requirements, and classroom characteristics, both in the K-12 and postsecondary classrooms (Abrams, Varier & McMillan, 2013; McMillan, 2003; Yorke, 2003).

Framing their empirical work using a sociocultural approach, Pryor and Crossouard (2008) theorized formative assessment as a "discursive social practice" (p. 1). They reviewed their empirical work on formative assessment, wherein the teacher takes on the multiple identities/roles of a content expert who models and provides criteria for performance, an assessor who carefully uses evaluative language to interact with the student, and a learner who narrates and critiques the process of learning. Empirical research with an explicit sociocultural framework has been conducted by researchers in the form of smaller, exploratory qualitative case

studies (e.g., Willis, 2011), as well as larger studies (Pryor & Crossouard, 2008). This line of research is promising because by shifting the focus to teacher-student learning interactions, it successfully removes the notion of formative assessment as simply a means of measurement.

The Study of Formative Assessment Using Sociocultural Theory

Willis (2011) studied teacher-student interactions during formative assessment in three middle school classrooms in Australia. Preferring instead the term “assessment for learning,” Willis viewed formative assessment as “culturally situated patterns of participation that allow students to develop a sense of belonging as an insider in the practice, while developing an identity of an autonomous learner” (p. 402). By participating in formative assessment with teachers, peers, and self, the learner has the opportunity to reflect on developing expertise in the process of learning. The study reported findings that represented quintessential features of formative assessment; for example, all teachers shared criteria for success with their students, engaged in activities to elicit evidence on where students are performance-wise, provided ample feedback, and cared about what students did with the feedback. However, Willis’ observations and focus group findings presented a classroom culture wherein students became central to the formative assessment process; the students renegotiated their previous beliefs about their role in assessment according to the expectation their teachers conveyed through students’ participation. Where there was a match between classroom expectations and student beliefs, there were positive teacher-student relationships, and vice versa.

In a research study conducted in Scotland, Crossouard (2011) used Vygotsky’s cultural historical activity theory in combination with Sadler’s (1989) formative assessment framework. This resulted in an assessment design that treated teacher-student dialogue as a necessary component in the formative assessment process. Scotland’s educational context defined

development as the process that makes students “successful learners, confident individuals, responsible citizens, and effective contributors to society” (p. 62), and not just masters of academic content. Crossouard’s study was an in-depth case study that examined formative assessment practices in two elementary classrooms that served socially and economically disadvantaged children, in which participating teachers reported valuing formative assessment. The classroom activities reported by the researcher reflected several essential characteristics conducive to formative assessment; however, Crossouard found that the language teachers used was largely behaviorist, and focused on positive reinforcement rather than fostering a discourse about learning. For example, teachers encouraged students to participate by asking them to provide observable behaviors to demonstrate their level of learning; this led to student demonstrations of learning (by reporting the right or wrong answer), rather than discussions about cognitive processes or student thinking. As long as student identities in the classroom remain in the periphery of the learning process as a result of teacher expectations, the true purpose of formative assessment remains unfulfilled. In this case, teachers maintained authoritative expectations of students instead of opening up the learning environment for negotiation. This study serves as an example of how the sociocultural lens can illustrate problems in current formative assessment practices.

Similar studies have also been conducted in several other contexts. For example, Pryor and Crossouard (2010) examined formative assessment in relation to student identity in a postgraduate context. Collecting data on teacher responses to student work, and student responses to teacher feedback, they posited that learning is not a discrete activity, but a continuous dialogue in which teacher and learner identities define the practices and nature of learning. Another example includes the study of formative assessment in English classrooms in

Malaysian primary schools (Sardareh & Saab, 2012). In this study, Sardareh and Saab observed and conducted interviews with three teachers to understand their use of formative assessment. Based on their findings, the researchers developed a sociocultural model of formative assessment that emphasized the explicit sharing of criteria, use of collaboration and authentic learning activities, and encouragement of reflection through feedback and discourse, as well as active learning.

Pryor and Crossouard (2008) theorized formative assessment using a sociocultural framework. They found that assessments in classroom settings could be classified as convergent or divergent. Convergent assessments were those assessments that teachers developed and used to see if learning had taken place. These assessments included closed- and open-ended questions with definitive correct responses. The teacher's position in such an assessment task was authoritative, and feedback primarily included corrections of student errors. Alternatively, divergent assessments were open tasks that aimed to get at what the students have learned. In divergent assessments, the teachers acted as collaborators, and feedback tended to be descriptive and served as an important opportunity to make metacognitive reflection of learning explicit. However, Pryor and Crossouard noted that convergent assessment should not be viewed as bad or counterproductive to learning, as this type of assessment serves the important function of determining if students have learned in classroom contexts.

Pryor and Crossouard (2008) conceptualized formative assessment practice as involving teacher-student interactions, where teachers engaged in both "helping" questions and "testing" questions, in addition to observation of the learning process and outcome. Through feedback and judgment, they mediated students' understanding in order to help them meet learning goals. Therefore, formative assessment is not a discrete practice apart from learning, but rather includes

ongoing teacher-student dialogue. Similarly, formative assessment practice in the use of formal assessments is not separate from learning, and includes dialogue closely connected to the learning process in the form of teacher feedback and student uptake.

Pryor and Crossouard's (2008) study uses the principles of formative assessment theory to understand formative assessment practices in ESL classrooms. Formative assessment is conceptualized as both feedback patterns that occur in daily classroom interactions, as well as the use of formal assessments to inform instruction and support student learning. The following section describes empirical literature on formative assessment in higher education learning contexts, particularly, in the ESL classroom.

The Study of Formative Assessment in Higher Education ESL/EFL Contexts

In the context of higher education, as a result of institutional goals and expectations for teaching staff, and the focus on general outcomes, such as graduation and retention as measures of success, formative assessment is not prioritized in instructional contexts. These contextual expectations, in turn, affect faculty members' instructional practices. Furthermore, contextual expectations and goals differ substantially according to discipline, course, and academic level. These differences, and the how they play out in the assessment practices of teachers, are not well understood (Gibbs & Simpson, 2004). Therefore, generalizing the use and practice of formative assessment to all of higher education is difficult, and perhaps even undesirable.

As noted previously in this review of the literature, the vast body of empirical and theoretical literature is informed by K-12-based formative assessment. The emphasis on K-12 contexts has contributed to several issues in our understanding of formative assessment, including the tendency to oversimplify formative assessment as a means to improve performance on summative assessments (Abrams, 2007). In a widely cited article on formative assessment in

higher education, Yorke (2003) stated the need to consider intellectual development and the psychology of feedback in theorizing formative assessment.

The following section describes relevant studies on formative assessment conducted in postsecondary ESL classrooms. These studies include: 1) studies on feedback interventions; 2) teacher feedback practices; 3) student perceptions of feedback and assessment; and 4) formative assessment practices, including classroom-based and online formative assessment.

Studies on Feedback Interventions

Ashwell (2000) conducted a quasi-experimental study where English language learners received one of four feedback conditions on three drafts of an essay. Using a process-oriented writing approach, learners received either no feedback, feedback on content followed by form (e.g., grammar, mechanics), feedback on form followed by content, or a mix of the two. They found that the group that received no feedback was significantly different from the other three feedback condition groups; in each of the three feedback conditions, there was a reduction in the number of grammatical errors and an improvement in content scores. In most cases, the three conditions showed significant changes in scores between the first and the second, as well as the second and third, drafts. The researchers concluded that the recommended approach in process writing, which focuses on content-level feedback followed by form-related feedback, was as useful as mixing feedback and following a reverse pattern of form first and then content. However, it is important to note that in the latter case, where feedback focused on form followed by content, students made considerably fewer grammatical errors in the second draft (after feedback on form), but showed an increase in the number of grammatical errors in the third draft (after feedback on content). This finding questions the efficacy of form-related feedback.

Corrective feedback may lead to changes in revisions, but is not necessarily indicative of student learning resulting from feedback.

A number of studies have been conducted on the role of written feedback or feedback on performance on writing tasks. Guenette (2007) examined studies published in the last 20 years on feedback in writing to understand the extent to which knowledge on best practices regarding feedback may be generalized. By analyzing quantitative studies that predominantly included a control group that received no feedback in naturally occurring groups (classrooms), he found that the studies were too different in their eliciting tasks and instructional practices, making a comparison of them inappropriate. Furthermore, feedback interventions often examined the changes made to drafts as an outcome measure that is influenced by the nature and amount of corrective feedback provided by the teacher, and not necessarily student learning from feedback. The study conclusion urged for more descriptive studies that consider the whole context “in and out of the classroom” (p. 51) and individual motivational characteristics while studying the effect of feedback on ESL student outcomes.

Teachers’ Feedback Practices

In a study of ESL teachers, Bailey and Garner (2011) conducted 48 interviews with faculty members at a British university to examine their experiences providing written feedback on student work. They found that faculty members gave feedback to help students make corrections and improvements to their work, and enhance student motivation, but they also provided it as a response to university policies on providing feedback and justifying grades. Participants commented on how they thought students used feedback, and they presented students who used feedback as motivated and more able, while students who did not use feedback were portrayed as weak and indifferent. A major theme that emerged from Bailey and

Garner's study was that faculty members believed that the reason students did not use feedback was because they did not understand it. Within the context of the study, written feedback on assessments was provided on a separate form, which made it difficult for students to connect the feedback to specific aspects of their performance or make meaningful changes based on the feedback provided. The researchers concluded that the efficacy of feedback based on participant accounts was, at best, ambivalent. Whereas the faculty members perceived feedback to be useful for students to improve their performance, they did not know how students understood and utilized feedback. In addition, institutional factors played a role in how feedback was provided. A final limitation of the study is that it did not cite participating teachers' class sizes, which may also affect their feedback beliefs and practices.

Similar to Bailey and Garner (2011), Ferris, Pezone, Tade and Tinti (1997) studied more than 1000 comments provided by a teacher to advanced ESL students on 111 first drafts. They analyzed the nature of teacher feedback with regard to pragmatic (who gets feedback, what kind of feedback is based on tasks, etc.) and linguistic form (asking for information, making suggestions, correcting grammar, etc.). They also sought to determine if teacher feedback varied based on student ability, assignment types, and the time in the semester. The student population in this study was different from typical English as Foreign Language (EFL) students, however, in that most students went to high school in the United States and were residents, even though they came from a variety of ethnic backgrounds. Ferris et al. analyzed the teacher's comments and categorized them based on student ability, assignment type, and semester; they also noted whether the feedback was positive, directive, or related to grammar. They found significant differences between the frequencies of comments made for weak students as compared to strong students, and also differences early in the semester versus mid and late semester. The researchers

concluded that the rationale informing teachers' decision making regarding the quantity and nature of feedback is ultimately pragmatic, and varies according to the individual learner and the time of semester, two variables that remain understudied in current research on ESL feedback practices.

In addition to examining teachers' feedback practices, some studies, such as Li and Barnard's (2011), expanded the scope of their study so as to include tutors' feedback practices. Li and Barnard conducted an in-depth case study of ESL tutors whose main task was to provide feedback to undergraduate students on their writing. The study investigated the tutors' beliefs and practices regarding feedback and grading. Using multiple data collection procedures, including questionnaires, interviews, think-aloud procedures, stimulated recall, and focus groups, the researchers identified three reasons for incorporating feedback into instruction: 1) the need to improve writing in the future, 2) the need to provide positive feedback along with negative feedback, and 3) the need to justify grades. The tutors in this study were students themselves, and they used their own experiences receiving feedback to inform their practice. The researchers found tutors added smiley faces and positive comments like "well done," but overall, there were several differences in the quantity and nature of feedback provided, ranging from short phrases like "great work" to long corrective paragraphs. Several participants reported rereading assignments to check whether their grading was appropriate, and oftentimes made changes to the grade upon rereading. The researchers claimed that although tutors mentioned improving students' writing as a major motivation for feedback, in practice, several factors were actually at play. One such factor, the need to justify grades by providing feedback, has interesting implications for future research. Research in the realm of formative feedback may need to explore the magnitude and prevalence of alternate rationales guiding feedback practice.

In a study that explored the rationale behind teachers' ratings of student essays in a non-traditional ESL setting, Cumming, Kantor and Powers (2002) analyzed the think-alouds of seven ESL faculty members who scored a set of Test of English as a Foreign Language (TOEFL) exam essays. The researchers found that raters referenced other essays in the set they were reviewing when making decisions about scores, which points to a potential norm-referenced decision making process. In deciding upon a score, they attempted to balance positive and negative aspects of content as well as language and grammar. They found that teachers shared similar views as to what constitutes effective writing, and they also found that teachers tended to focus on the rhetoric and ideas of high-performing essays more than they did for weakly written essays. They reported bringing their previous experience as ESL teachers to bear while rating the essays. The researchers also noted that more experienced raters verbalized their thoughts more often than newer raters. There were also differences among the raters who were native speakers of English in comparison to raters who spoke English as a second language. The latter took longer, and engaged with the essays more than the native English speakers; they also verbalized their decisions more than native speakers. The researchers noted that the background of each may play a role in their scoring practices, with native speakers being more concerned with literary quality of writing, and non-native speakers more with the pragmatic aims of writing. Although not directly reflective of classroom assessment, this study implies that teacher characteristics affect their assessment practices.

Still on the subject of English language learning and instruction, Aljaafreh and Lantolf (1994) were interested in the transition of English language development, particularly writing from an intermental (social) to an intramental (individual) plane in ESL teaching and learning contexts. In a longitudinal study, they analyzed teacher-student interactions in written work that

operated within the zone of proximal development. The students featured in the study were intermediate learners of English, enrolled in an eight-week course at a university in the United States. The study took place outside of the class, where the researchers served as tutors to provide feedback on writing. They looked for two outcomes, the first of which was a product oriented-outcome, which involved reduction of errors and improvement in the written work. The second outcome was more implicit, and sought to see if the learner had moved from other-regulation to self-regulation levels, where he/she could identify errors previously pointed out by the tutor. Aljaafreh and Lantolf found clear indications that the tutors and learners were operating from the ZPD, and that learning was occurring in the moment. The study found differential levels of feedback negotiation between learners, and it provided an explicit mediational framework using a sociocultural perspective. Findings from this study are limited on account of specific language-related interactions, i.e., error corrections unique to a tutor-learner environment that do not generalize to a typical ESL classroom setting. Still, the findings point to the potential of examining feedback practices using an sociocultural framework.

Hyland and Hyland (2001) investigated the kind and quality of feedback provided by teachers in a more traditional EFL setting. They conducted an in-depth case study of three EFL teachers in undergraduate and graduate writing classes in Japan. They used multiple methods, including written composition and feedback data, think-alouds while teachers were writing feedback on student work, and student interviews about their reactions to triangulate feedback practices. They found that of 500 instances, 40% of feedback was praise-related. Teachers also gave more praise-related feedback than negative or suggestion-based feedback, which was even more prevalent in final drafts. Though, teachers differed in their feedback practices. For example, one teacher found it difficult to provide negative feedback without couching it in

positive feedback. In another case, a teacher tended to provide comments at the end of the document, demonstrating a preference for holistic, rather than specific, feedback.

Hyland and Hyland (2001) also noted several instances where teachers referred back to previous advice and conversations with students in giving feedback, thus providing evidence of a dialogic process. Often, teachers paired two types of feedback, such as positive and negative, negative and suggestive, etc. The researchers argued that this could highlight the teacher-student relationship as a reason why teachers found the need to mitigate their feedback. Teachers mitigated their comments for several reasons, including the desire to avoid hurting and discouraging students by supplementing negative feedback with positive feedback. As teachers negotiated the dual roles of reader and authoritative assessor, they tried to retain student ideas in spite of criticism by using language like "try to" and "I suggest." Student interviews featured in this study revealed that students found positive comments to be motivating; though, students nonetheless varied in their reactions to positive feedback, with views ranging from motivating to useless. At the same time, they recognized teachers' use of words like "but" following positive comments as a way to construct criticism.

A major theme in student interviews was the difficulty students experienced in understanding feedback. Subtle comments from teachers, especially indirect criticisms, were found to be confusing and were less informative. In their think-alouds, teachers expressed awareness as to whether students would be able to follow their feedback if it were written in an indirect manner. As an example of the indirect feedback teachers sometimes provided, one teacher was looking for the definition of a term in the student's introductory paragraph, and commented that the student had introduced many terms without defining any of them, especially one important term; she wrote, "I should like to know more about macroscopic" (Hyland &

Hyland, 2001, p. 203). Furthermore, the student failed to include definitions in the final draft, to which the teacher responded, "Still hasn't changed." When asked about why the first comment was left unaddressed, the student responded that he did not understand it as a form of feedback requiring action on his part. Instead, he interpreted the comment as indicating that he did not understand the term very well. In another instance, a teacher commented on a student's repetition of ideas, directing the student's attention to it with an indirect comment: "This second section seemed like a summary of the previous information" (p. 203). Once again, this student likewise misinterpreted the comment as a result of its relative indirection, but this time, the student took the comment to be positive. Such feedback interactions point to the difficulty EFL students may have in discerning implicit feedback due to their limited proficiency.

Student Perceptions of Formative Feedback

Hedgcock and Lefkowitz (1994) examined student reactions to feedback in the context of the increasingly popular research on feedback intervention. They sought to understand the connections between ESL and foreign language students' reactions to feedback and changes in their perception of text quality, as well as whether the students differed in their responses to feedback. The researchers administered an in-depth survey on teachers' feedback practices to 247 beginner-level foreign language learners. Hedgcock and Lefkowitz found that ESL students' perceptions of their instructors' feedback practices regarding the type and focus of feedback was consistent with their expectations for feedback from instructors. The students indicated that feedback on content was a priority for them, even though they still benefited from comments on grammar corrections, organization, and style. Generally, students held positive attitudes toward writing, and rated themselves highly on a self-assessment item that asked them if they were highly competent ESL writers in relation to their peers. In contrast, foreign language students

(non-ESL students) reported feedback on grammar, vocabulary, and mechanics as most important. This was consistent with their expectations for instructors regarding feedback on their drafts. Both groups indicated that the most meaningful changes in their writing resulted from instructors providing them with written comments, and meeting with them on an individual basis.

In another study, Brice (1995) video-recorded three students' reactions to teacher commentary on their essay drafts. The primary objective of the study was to examine the extent to which students understood, liked, and found beneficial teachers' feedback on their writing. Classifying feedback comments according to type, explicit and implicit, Brice found that students appreciated teachers' corrective feedback and improvement suggestions. They struggled, however, with implicit, unclear comments that did not clearly indicate next steps, especially as it concerned grammar and vocabulary.

In addition to Brice's study, other studies also speak to the difficulties EFL students experience in interpreting instructor feedback. In a study of oral feedback provided during classroom interactions, Panova and Lyster (2002) found that recasts (repeating the error to the learner) and translation comprised the majority of feedback provided to EFL students in oral communication classrooms. Metalinguistic feedback, which refers to providing clues intended to trigger thinking in students to repair their language use, by contrast, occurred less frequently. Overall, of the 412 feedback opportunities, only 16% resulted in learner acceptance of feedback and subsequent correction of error. While Panova and Lyster's study explored types of oral feedback, the low rate of successful error correction may be attributed to implicit correction techniques, or minimum scaffolding to help the learner detect errors and consequently make corrections.

Higgins, Hartley and Skleton's (2002) study also shared the goal of exploring student perceptions toward feedback. The researchers conducted a mixed methods study in the UK, using questionnaires and interviews. They found that 40% of the students surveyed reported having trouble understanding feedback—either because of handwriting, complex academic language, or the vague nature of feedback. They also tended to be negative about feedback that they perceived to be less useful. A majority of them (81%) looked over written feedback, and spent less than 15 minutes going over the feedback they received, yet, about the same number of students disagreed that feedback was not useful. They viewed feedback as a means to identify errors to help inform future writing, and thus improve their grades. The researchers discussed several implications of the study, including utilizing peer assessments to improve student understanding of criteria, encouraging teacher-student dialogue on learning expectations, and establishing criteria to help students understand feedback in relation to said criteria. They also noted that students require explanations to aid them in evaluating their mistakes to inform future performance.

Against a backdrop of feedback related case studies in the ESL literature, there are studies that have examined feedback practice using an explicit formative assessment framework. The following section elaborates on the context, issues, and empirical investigations surrounding formative assessment in ESL settings.

Formative Assessment Practices in ESL Contexts

Cumming (2009) discussed how EFL/ESL contexts can use classroom assessments. She noted there are increasing commonalities across English language curricula (for example, TESOL standards, Common European Framework of Reference), which have influenced local (institutional) expectations of proficiency. This relative standardization across curricula creates

unique, and often misunderstood, issues concerning teachers' practices in the classroom. Further, assessments have taken on a central role in English language teaching and learning, as well as teacher practices with regard to classroom assessment and other fine-grained assessments; however, despite their central role, assessments in English language contexts have not been sufficiently examined. Cumming concluded that there is a need for better understanding of the context, teacher-student interactions, and applications of assessments to further student learning in EFL classrooms. As noted in the previous sections, a substantial portion of related research has been conducted on EFL or second language acquisition (Clark & Clark, 2008; Leki, Cumming, & Silva, 2008), which could be better studied by using the formative assessment lens.

Recognizing this need for the study of formative assessment in the EFL classroom, Davison and Leung (2009) pitted teacher-based assessments against large scale norm-referenced assessments to present several important issues plaguing formative assessment in the language learning context. The use of meaningful assessments that foster reflection on the part of teachers and students is an important aspect of teacher-based formative assessment. In language learning, this may be conceptualized as authentic assessments that use language in realistic settings. A major problem with such assessments in EFL contexts is the confusion surrounding what is being assessed; in addition to language, assessments reflecting authentic use of language (such as conversing in a social gathering, asking for directions, etc.) also measure learners' knowledge of culture and social skills. Although most EFL contexts aim to provide learners with sociocultural competence in addition to language proficiency, this raises an important question about the validity and reliability of such assessments as reflections of learning.

The lack of a solid theory that views learning as informing the practice of formative assessment may trivialize teacher-based assessment use. The use of sociocultural concepts, such as mediation and interaction, to comprise a theory of formative assessment is integral to EFL teaching and learning. When considering teacher-based assessments, there is no guidance for teachers as to when they should support learners in their attempts to understand and express language, and when they are to stand back and evaluate learners' level of learning. Formative assessments provide ecological validity by gathering information on what students know, and require contexts that encourage the use of formative assessment to demonstrate trust and provide autonomy to teachers.

Ultimately, students must demonstrate learning in large-scale assessments. An important question then is regarding the alignment of classroom-based assessments with standardized assessments and other widely used measures of language competency, like the TOEFL. Hamp-Lyons (1994) noted that EFL/ESL contexts have entrance and exit assessments based on which key educational decisions are made. These assessments are not useful for teachers, who need ways to provide ongoing support and feedback to students. Hamp-Lyons, however, provided several ways to incorporate formative assessment into instruction, including the use of self-assessment, peer assessment, process-oriented writing approaches, and portfolio-based assessments, all of which are designed to foster evaluative skills in students to help them become aware of their learning and progress.

Cheng, Rogers, and Wang (2007) compared classroom assessment practices in three EFL contexts: Canada, Hong Kong and China. They interviewed 74 instructors in seven or eight universities in each country. They found that context goals and the purpose of language learning influenced instructors' assessment practices. The researchers discovered that in China, a primary

goal of the EFL class was to prepare students for a standardized English assessment that would enable students' entry into English medium universities. Class sizes were large (about 31 students), and teachers used summative assessments and examinations that accounted for 80% of students' final grades. In contrast, Canadian and Hong Kong classrooms were smaller (10–15 students), and the purpose of the class was to make students proficient in English. In these settings, 75% of assessments included student work composed throughout the course, as well as portfolios, presentations, and journals. These assessments served a formative purpose; they were used to acquire insight into student understanding and progress, and to identify weaknesses. In China, on the other hand, classroom assessments were used to formally record student progress and prepare them for the standardized college English test.

As these studies demonstrate, recent empirical investigations have emerged from countries other than the United States with an explicit focus on formative assessment practices. While a majority of the studies have been qualitative or mixed method case studies, Birjandi and Tamjid (2012) conducted a quasi-experimental five-group intervention on the effect of peer and teacher assessment on Iranian English learners. They studied the impact of self-assessment, peer-assessment, and teacher assessment on writing performance. During a semester-long intervention, they randomly assigned 157 intermediate-level EFL learners to four treatments, and designated one control group across five predetermined writing classrooms. They found that all groups significantly improved their performance. However, compared to the control group, the groups that engaged in a combination of self-assessment and teacher assessment (effect size = .09), as well as peer assessment and teacher assessment (effect size = .19), exhibited a significant difference in terms of performance. There was no improvement noted for the group engaged in a combination of journal writing or self-assessment and peer assessment with no teacher

assessment. Findings indicate the importance of individual characteristics, such as overt self-assessment, and contextual factors, like peer and teacher feedback, in learning writing.

In a case study of two teachers in tertiary EFL classrooms, Chen, May, Klenowski, and Kettle (2013a) conducted teacher interviews, student interviews, and classroom observations to investigate how Chinese policy reforms regarding formative assessment were implemented in classrooms. The researchers posited that the foundation of Chinese education is based on testing, and recent legislative reforms requiring increased self and peer assessment in classrooms contradict traditional views of educational contexts. This testing culture permeated teacher beliefs and practices, as well as student beliefs about learning. They found that both teachers were confident about their ability to create meaningful assessments, and demonstrated autonomy and individual preferences with respect to the use of assessments and feedback practices. However, their feedback practices differed, as one teacher focused on task-based corrective feedback, with her practice primarily oriented towards preparing for tests. Conversely, the other teacher resisted student pressure to emphasize testing and deemphasize peer feedback activities. Chen et al.'s study points to important contextual and cultural factors at play in English language classrooms, including student beliefs about learning and assessment and teacher beliefs and pedagogical practices, all of which interact with institutional expectations. Additionally, progress in policymaking did not automatically trigger change, and participants showed strong beliefs in support of the traditional teacher-centered classroom.

In another study, Chen, Kettle, Klenowski and May (2013b) examined how local contexts enacted formative assessment mandated by national education policy reform. Moving away from a complete emphasis on one final assessment, or a single score on a mandatory standardized assessment called the College English Test, this new policy required making room

for process-orientated language instruction and assessment, in addition to peer and self-assessment. However, institutions translated these national mandates into local policies that focused on classroom participation, quizzes, and assignments, all of which are graded and count toward the student's final grade. Although classified as formative assessment, this institutional move emphasized multiple sources of grading (as opposed to just one), and greater incorporation of teacher judgments in grading. Chen and colleagues conducted administrator and teacher interviews, in conjunction with focus groups, to identify the sociocultural contexts that defined the practice of formative assessment in the two universities. They found that the two participating universities differed in their enactment of formative assessment: While one university allotted 60% of the course grade to process-oriented assessment, the second university only allotted 10%. Although there were key socioeconomic, geographic, and other demographic differences between the two universities, the university that assigned 60% process-oriented grading demonstrated a sense of trust in teachers' judgments, contrary to the other university. They found that teachers responses reflected a high degree of acceptance of policy-related changes, but also described having flexibility with the type of assessments they could administer within the limitations imposed by the process-based approach.

While the above studies illustrate the tension between contextual and classroom expectations in relation to formative assessment, there are important factors between teachers and students that affect formative assessment practices. Hernandez (2012) investigated how continuous assessment (as opposed to interim and final assessments only) supported learning in higher education contexts in Ireland. Hernandez gathered data from 138 undergraduate students and 41 professors of Hispanic Studies from seven universities. Using multiple data collection methods, including student interviews and teacher questionnaires, and follow-up interviews with

a subset of teachers, the study found mixed results regarding the role of assessment in relation to feedback and grades. Students reported that grades were an important part of assessment, whereas professors placed less emphasis on assessment as primarily functioning to assign grades. Students also expressed dissatisfaction with respect to the feedback they received, viewing it as too little, vague, or delayed to help them make substantial changes. Professors, for the most part, expected students to act on the feedback, but only some provided opportunities for students to act on the feedback to fulfill this requirement. Some students noted that too many graded assignments, as seen in continuous assessment, could lead to increased anxiety about assessment. On the other hand, professors indicated that attaching grades makes students care about the assessment, yet once the grade is obtained, students are not sufficiently motivated to improve their learning.

Synthesis

The study of instructional and assessment practices grounded in the theory and principles of formative assessment is not well developed in postsecondary EFL/ESL teaching and learning contexts. The purpose of the previous section was to construct a picture of feedback practices that can be expected in ESL settings. As several researchers have pointed out, there is no clear consensus on how learning occurs through feedback-related interactions, and there is much room for inquiry into how student perceptions are related to teacher feedback practices. Literature on global assessment practices also points to the influence of institutional and government policies on the use of assessments. As demonstrated by the extensive literature on ESL writing, there are several teacher characteristics, student characteristics, and contextual variables that play a role in the provision and reception of feedback. It follows that there is a need to better understand formative assessment practices in relation to contextual influences, teacher characteristics, and

student characteristics, such as their metacognitive beliefs, motivation, and conceptions of learning and assessment.

The following section introduces the concept of metacognition, and the vast amount of literature that has led to important understandings in learners' cognitive processes. This research provides insight into the practical application of metacognitive measures in the classroom. Andrade (2013), for example, presented self-regulated learning research as a context in which we should consider classroom assessment. Self-regulated learning has implications for formative assessment in that monitoring performance, evaluating one's own performance, and incorporating external feedback are not only important features of self-regulated learning behavior, but also essential goals of formative assessment. By integrating students' perceptions of their learning with formative feedback, teachers and students can actively work towards achieving learning goals. Students can provide useful information to teachers on their current learning by reporting their thought processes on assessment tasks. Teachers can gauge whether student reports on their knowledge and confidence level regarding assessment content are consistent with their performance. Inconsistencies may be indicative of misconceptions at the individual level, even at the group or class level.

The Role of Metacognition in Formative Assessment

There is increased interest in using assessments as a way to provide information on the learner's cognitive processes. In the last decade, a vast amount of research has been conducted on the application of our knowledge of student self-regulated learning processes to enhance learning; for example, one study investigated the use of computer-based programs (called intelligent tutors) as a means of supporting students' self-regulatory skills as they engage in subject-based performance tasks (Greene & Azevedo, 2010). Learners engage in self-regulated

learning when they set goals, plan actions, activate relevant cognitive skills like memory and learning strategies, and control their non-cognitive skills like motivation during learning (Pintrich, 2000).

Self-regulated learning is a complex construct that involves several factors. From a sociocognitive perspective, self-regulated learning is the process of initiating goal-directed behavior to acquire knowledge through the appropriate use of strategies. Successful learners demonstrate high self-regulatory skills, and there is a strong positive relationship between one's self-regulated learning ability and his or her academic achievement (Zimmerman, 1990). Self-regulated learning theory came out of Bandura's social learning theory and theory of reciprocal determinism. *Reciprocal determinism* refers to the triadic relationship between the personal, behavioral, and environmental factors involved in learning and development. Depending on the task or context, one factor, or determinant, may be more influential than the other. For example, in a context where a student receives instruction that modeled procedural steps for solving a problem, she may not use her knowledge from past experience, or her own strategy to problem solve. In a context where she is encouraged to problem solve on her own, the student activates more personal and behavioral factors (Zimmerman, 1989; Zimmerman & Timothy, 2004).

There are two models of self-regulated learning that are widely accepted and studied (Winne, 2010; Cleary & Zimmerman, 2004; Winne & Hadwin, 2000). In the first model, self-regulated learning is a cyclical process that involves planning or forethought before taking on a task, monitoring and making adjustments during learning, and lastly, evaluating the resulting performance, known as the self-evaluation phase (Cleary & Zimmerman, 2004). Here, self-regulated learning takes on stable characteristics understood as a general self-regulated learning aptitude. The other model of self-regulated learning describes it as an event involving four

phases: (1) defining a learning task, (2) setting goals and understanding criteria, (3) activating and enacting behaviors and strategies to learn or achieve goals, and (4) finally, evaluating the three phases to inform future learning (Winne & Hadwin, 2000). This second model is more concerned with learner decision making while engaged in a task. Although the two models, in essence, capture the planning, monitoring, and evaluating that individuals engage in while learning, the first model, the Zimmerman model, conceptualizes self-regulated learning as a trait and skill that develops over time. This does not imply domain generality of self-regulated learning, but that there is a global skill and aptitude learners acquire and use over time. By contrast, the second model is concerned with actual behaviors exhibited by learners during specific learning events, such as learning a new concept, or preparing for a class test. However, both models are concerned with behaviors associated with the learning process.

Several factors contribute to self-regulatory behaviors in students, including prior knowledge, motivational beliefs, and metacognition (Sperling, Howard, Staley, & DuBois, 2004). Prior knowledge includes familiarity with the learning task and its associated content, as well as knowledge of past performances. Motivational beliefs are student self-beliefs regarding their interest, ability to complete learning tasks successfully (self-efficacy), and their purpose for learning (goal orientation). Metacognition refers to learners' knowledge about their cognitive ability, including an awareness of the skills they possess for learning, and their ability to activate, use, and modify said skills. Metacognition and motivation have been studied extensively in the field of education and psychology, and both have been shown to be important contributors to learning, as well as self-regulation (Veenman, 2011; McCormick, 2003; Schneider, 2008). However, metacognition, as a broader term encompassing individuals' awareness and control of their cognitive processes in all types of learning, is a vast area of research itself. Metacognition

is often confounded with self-regulated learning, as researchers have used the terms interchangeably (Dinsmore, Alexander, & Loughlin, 2008). The following section provides a brief overview of the relationship between self-regulated learning and metacognition with the purpose of delineating which construct is more closely associated with formative assessment in classrooms.

The Relationship Between Metacognition and Self-regulated Learning

In her introduction to a special issue on the relationship between metacognition, self-regulation, and self-regulated learning, Alexander (2008) discusses conceptual cross-fertilization, and the need to clarify and explore the conceptual boundaries between these three concepts. The literature often uses the three terms indiscriminately and interchangeably, and the terms have become so popular that no one feels the need to clarify their definitions. The term *self-regulation* in education literature often refers to self-regulated learning, whereas it refers to the general, global monitoring and evaluation of behavior in the context of psychotherapy and cognitive behavior management (Dinsmore et al., 2008).

Dinsmore et al. (2008) conducted an analysis of contemporary literature on metacognition, self-regulation, and self-regulated learning. Their analysis of the theoretical and conceptual underpinnings of each of these concepts is described below.

The concept of metacognition emerged from the work of Flavell (1979), who provided a framework for understanding “thinking about thinking,” which originally included four components: metacognitive knowledge, metacognitive experience, goals, and strategies (McCormick, 2003). Metacognition is primarily concerned with individuals’ reflections on their own cognitive structures, without the environment playing an explicit role in these reflections. Metacognitive theory has evolved since Flavell's original concept, and contemporary frameworks

describe metacognition in terms of two components. The first component, metacognitive knowledge, refers to learners' awareness of their own cognitive processes and cognitive processes in general. The second component, metacognitive control, refers to knowledge about how, when, and why learners activate relevant cognitive processes (McCormick, 2003; Schraw, 2009).

Self-regulation is an outcome of Bandura's theory of social learning. Dinsmore et al. (2008) labelled self-regulation as “exogenous constructivism” because it is the process and product of a person's interaction with his or her environment, mediated by behavior. The focus of self-regulation is not on individual cognitive processes that function and develop independent of the environment, but rather the result of one's interaction with the environment. Here, the environment plays a critical role in shaping internal processes. Self-regulated learning draws from both the environment, and the cognitive processes of self-regulation and metacognition. As self-regulation became more and more a part of education research, self-regulated learning became synonymous with self-regulation when applied to academic settings.

From their analysis, Dinsmore et al. (2008) determined that there are points of convergence and divergence between the terms self-regulation, self-regulated learning, and metacognition. At their conceptual core, these three terms are similar in that they are concerned with individuals' knowledge of their own cognitive processing, including the monitoring and controlling of their own cognitive repertoire through use of strategies. What distinguishes these terms from one another is the role of the environment in each concept. Metacognition is concerned with an individual's mind, and the development of thinking skills within an individual's cognitive repertoire. Self-regulation and self-regulated learning are comparatively more concerned with the environment, and its role in developing, facilitating, and constraining

an individual's self-regulatory processes. For example, the role of parental or teacher guidance in a student's self-regulated learning. Kaplan (2008) suggested that given their conceptual commonalities, metacognition and self-regulation may be construed as part of a broader concept underlying self-regulated action.

Kaplan (2008) argued that it is not desirable to define clear boundaries between metacognition and self-regulation because the two concepts are interrelated. Metacognition is the individuals' global awareness and control of cognitive processes in all learning and problem solving, in which academic learning represents only one part. As such, self-regulated learning may be subsumed under metacognition. Self-regulated learning, on the other hand, is the knowledge and control of cognitive processes in light of contextual expectations that facilitate or hinder learning academic constructs. Motivation and external influences are brought to bear in self-regulated learning, as are individuals' internal awareness and control of cognitive processes. As such, metacognition may be subsumed under self-regulated learning.

The solution provided by scholars of both constructs is for researchers to be explicit in their reasons for using metacognition or self-regulated learning (Schunk, 2008; Lajoie, 2008). Researchers should be aware of situations when one it may be more appropriate to employ one construct over the other. Scholars of metacognition and self-regulated learning also point to implications for the measurement of these constructs.

Sperling et al. (2004) conducted two studies with college students to test the relationship between metacognitive measures and self-regulated learning measures. In the first study, the researchers administered two metacognitive measures: the Metacognitive Awareness Inventory (MAI) (Schraw & Denison, 1994), a 52-item self-report questionnaire that asks participants to rate their metacognitive skills and strategy use, and the Learning Strategies Survey (Kardash &

Amlund, 1991), a 27-item questionnaire on covert and overt cognitive processes. The purpose of the study was to determine if metacognitive knowledge and metacognitive control were related. Sperling et al. found moderate yet significant correlations between the knowledge and control/regulation components of both measures. In the second study, the researchers investigated the relationship between metacognition and motivation, specifically, through the use of the Motivated Strategies Learning Questionnaire (MSLQ) (Garcia & Pintrich, 1995). The MSLQ is a measure of self-regulated learning strategies and motivation beliefs (goal orientation, task value, self-efficacy), and confidence judgments of test taking, which are prospective judgments of whether certain students are answering test questions correctly. They found a significant correlation between the accuracy of confidence judgments and metacognitive awareness and control. There were also significant positive correlations between measures of metacognition and learning strategies ($r = .60$). Furthermore, specific subscales of strategies like organization and time management were significantly related to knowledge and control, and effort regulation was significantly correlated with control of metacognitive processes. They found a positive correlation ($r = .37$) between accurate confidence judgments for correct answers and knowledge of cognition on the MAI. They also found students who were more accurate in their confidence judgments (low discrepancy between judgment and actual scores) showed high regulation of cognition on the MAI ($r = .42$).

The conceptual relationship between self-regulation and metacognition has implications for how these constructs are measured in research studies. Dinsmore et al. (2008) argued that it is important to clarify definitional and conceptual frameworks for self-regulation, metacognition, and self-regulated learning, as well as find more precise measures of the three constructs. They called for mixed methods studies to look at both individual processes and environmental

influences using a variety of sources, including prompting, self-reports, observations, and accuracy judgments.

In the majority of research, metacognitive and self-regulatory skills are measured using self-report questionnaires, most commonly, Likert scales (Schunk, 2008; Schellings & Hout-Wolters, 2011). Other methods have emerged to measure ongoing self-regulatory mechanisms used by learners, including think-aloud protocols where participants articulate their moment-by-moment thoughts as they engage in a task (Winne & Perry, 2000). Additionally, researchers who study self-regulated learning in computer-based learning environments measure self-regulated learning through trace data or computer logs that provide step-by-step details of the actions carried out by learners engaged in a learning activity (Greene & Azevedo, 2010; Lajoie, 2008).

A Review of the Theory and Applied Research of Metacognition

Flavell (1979) defined metacognition as the knowledge of one's own cognitive processes. Metacognition can be understood as being aware of one's own thinking and evaluating one's cognitive capacities as a way to monitor and modify actions in order to solve a problem (McCormick, 2003). This definition implies that metacognition is an internal process, maybe even automatic. It may also lead us to believe that metacognition is something we cannot control, and perhaps something that is not relevant to student learning or achievement. However, research since Flavell has shown that metacognition is both complex and critical to academic learning.

There are three main components of metacognition (Schneider, 2008; Veenman, Hout-Wolters, & Afflerbach, 2006): 1) metacognitive knowledge, which includes knowledge about personal strengths, weaknesses, tasks, strategies, and problem solving skills; 2) metacognitive monitoring, which involves paying attention to the process of learning as it occurs, recognizing both obstacles and facilitators of learning; and 3) metacognitive control, which results from

effective monitoring and refers to manipulating the process of learning to facilitate learning by using efficient strategies. Metacognitive knowledge can be classified into three types. The first type, declarative knowledge, refers to the knowledge that there are different types of tasks with varying difficulty, that there are personal factors like prior knowledge, confidence, and familiarity associated with the learning process, and that there are different strategies or approaches to problem solving. The second type, known as procedural knowledge, refers to knowing how to properly activate and apply a strategy in the course of learning. Finally, the third type of metacognitive knowledge, conditional knowledge, refers to knowing when and why one should use a particular strategy and approaching a task in a way that enables successful learning.

In education, metacognition is used to reference students' awareness of learning task characteristics, use of learning strategies, and interventions to improve metacognitive skills (McCormick, 2003). A vast amount of empirical work has indicated that metacognitive knowledge is an important factor in student learning and achievement (Pintrich, 2002). In fact, metacognition accounts for about 17% of unique variance in student performance—more than intelligence alone—and has an additional shared variance with intelligence of about 20% (Veenman et al., 2006). In a study of elementary students, Desoete (2008) found that metacognitive skillfulness and intelligence, together, accounted for at least 52.9% of variance in mathematics achievement. Knowledge and the ability to implement a variety of strategies are hallmark characteristics of expert learners. Additionally, knowledge of how experts approach a task has enabled researchers to build learning environments in ways that allow novice learners to acquire knowledge of strategies and skills to use while learning (e.g., Alexander, 2008; Lajoie & Azevedo, 2006). Research has found that metacognitive skills can be taught through instruction

and practice, and they have an impact on learning that is separate from intelligence (Winne, 2000). There is a significant amount of research conducted on interventions that foster skills in reading, writing, mathematics, and problem solving (McCormick, 2003). These interventions are often designed to provide explicit instruction in metacognition, which benefits a broad range of students (Palinscar & Brown, 1984; Veenman et al., 2006).

In describing how metacognitive tasks can be implemented in instruction, Weir (1998) recounted how she strategically embedded a metacognitive component into reading tasks, which included a blank line where students recorded their thinking as they read the passage. She also included a space where students had to record a predictive question. She reflected on student responses and noted that the metacognitive component created a classroom discourse about learning, produced higher-level student responses to questions, and encouraged students to feel in charge of their own learning.

Pintrich (2010) discussed the implications of metacognitive knowledge on teaching, learning, and assessment. He suggested that metacognitive knowledge needs to be made explicit in everyday instruction and interactions between teachers and students. He claimed that informal assessments that measure students' general metacognitive knowledge in the classroom can inform teachers' instructional decisions, for both content and metacognitive knowledge. Through "informal assessment conversations" (p. 224), teachers can also examine individual differences. It is likely that students who struggle with content knowledge may also indicate problems in metacognitive knowledge and strategies, which can be informative as teachers can adjust instruction or provide extra help. These strategies also enhance self-knowledge by making students' thought processes known.

Application of Metacognitive Judgments of Learning

Metacognitive judgments of learning refer to student perceptions of their level of preparedness, confidence in an academic skill or task, and the perceived difficulty of the learning material at hand (Dinsmore and Parkinson, 2013). The measurement of metacognition using self-report measures typically involves one or more of the following: confidence judgments, learning judgments, and/or difficulty/ease judgments (Dunlosky & Thiede, 2013; Hattie, 2013).

Confidence judgments refer to the extent to which the learner is confident that his or her response is correct. Learning or knowledge judgments may refer to if—and to what extent—a student learned the material being tested. Finally, difficulty judgments indicate the perceived level of difficulty of the material being tested. A learner who overestimates the accuracy of his performance demonstrates a poor metacognitive ability to monitor and evaluate learning and performance. This monitoring deficiency consequently may impact his control or use of appropriate strategies while learning. It is suggested, and studies have supported, that a domain-general metacognitive process may determine activation of domain-specific/relevant metacognitive skills based on task demands (McCormick, 2003; Veenman et al, 2006).

Metacognitive judgments of learning have been measured in a number of ways. For example, Metcalfe and Finn (2013) conducted several experiments with children in grades 3 and 5 to understand their ability to accurately judge their performance on tasks. They measured delayed judgments of learning (how well the children had learned on a scale of 1–100), and judgments of knowing (JOK), measured by whether or not the children knew the correct answer to items on a vocabulary test. Dinsmore and Parkinson (2013) similarly sought to measure metacognitive judgments in their study of undergraduate students. After reading a passage, students were asked to provide open-ended responses to their confidence judgments regarding the perceived difficulty of a test on the passage, along with responses to their confidence in the

accuracy of their answers. The researchers measured confidence judgments on a scale of 1–100, with 1 representing not confident and 100 representing confident. They found that students were mostly accurate in their metacognitive judgments and described multiple factors related to text characteristics, item characteristics, and prior knowledge to provide judgments. Another method that has been empirically well-studied is the Knowledge Monitoring Assessment (KMA), which is a dichotomous measure (yes/no) of students' reports on their prior learning while taking an assessment, such as reading or mathematics performance tasks (Tobias & Everson, 2009). The judgments are then analyzed by placing them in a matrix, categorizing them according to the correctness or incorrectness of each answer. In their discussion, Dinsmore and Parkinson (2013) noted that it is important to measure confidence judgments and task measures on a continuous scale to be able to determine the accuracy of judgments in relation to performance, and also suggested a need for further study on measuring judgments of learning.

Schraw (2009) provided five ways to analyze metacognitive judgments in relation to student performance (also known as calibration). The first means of analyzing metacognitive judgments is absolute accuracy, which refers to the precision of a student's judgment of learning and performance, where a discrepancy points to an under- or over-estimation of performance. The second, relative accuracy, refers to the relationship between students' judgments of learning and performance. Relative accuracy is usually measured by correlation coefficients or gamma coefficients, with discrepancies being indicated by the positive or negative sign. Bias, the third method for analyzing metacognitive judgments, refers to the degree to which the individual is over- or under- confident when making a confidence judgment. The index provides a measure of under- or over-confidence as a standard deviation or distance from 0, where 0 refers to accurate judgment, and hence is a measure of direction and magnitude of judgment accuracy. The next

method, scatter index, refers to the variance in judgments of correct and incorrect responses, with an index close to 0 indicating that judgments vary in a similar way for correct and incorrect responses, and are therefore, a reliable measure of variability. Discrimination index, the fifth and final method, is calculated to represent the degree to which an individual's judgments are different for correct and incorrect responses. Positive discrimination indicates the student is more confident in correct answers, and negative indicates greater confidence in incorrect answers. This index provides an in-depth measure of students' metacognitive monitoring skills.

Schraw (2009) noted important implications for researchers using metacognitive judgments of learning. One, it is important to understand the construct being measured in the study to inform the use of the calibration measure. He recommended the use of multiple outcome measures where appropriate. The grain size of the judgments, item-by-item versus one global judgment, influences the construct and interpretation of calibration measures. Two, the timing of the administration of judgments is important. Research has found delayed judgments of learning to be more accurate and reliable when compared to judgments provided immediately after a learning task (Hattie, 2013).

In his conclusion, Schraw (2009) addressed the need for more research connecting the calibration measures provided above to students' self-regulated learning repertoires or future learning outcomes. Hattie (2013) put calibration, or the accuracy between a student's judgment of his/her learning and actual performance, in the context of formative assessment. The issue of calibration is important for teachers because "by ignoring student's beliefs concerning their confidence and accuracy, we are ignoring a major precursor to their learning" (p. 62). Student judgments about their learning are indicative of their metacognitive awareness, or lack thereof, and monitoring behaviors that are important for self-regulated learning. Hattie discussed the role

of overconfidence in students, and the importance of teachers being aware of such miscalculations in order to make instructional adjustments. Overconfidence with regard to learning when the student has, in fact, not learned information, can hinder learning. Overconfidence may also point to the role of prior knowledge, which can play a powerful part in an assessment situation (Hattie, 2013). Students may revert to old strategies if they find it difficult to implement new learning, even if they show overconfidence in their judgments. Metacognitive judgments can provide important additional information that can help teachers identify instructional strategies that target various levels of learning and metacognitive beliefs, such as judgments related to ability and performance.

A major issue to be noted in the review of the literature on metacognition is that researchers use tasks that mimic real-world academic content. However, these studies are, in essence, empirical studies conducted in labs or experimental settings. Yet, the researchers discuss the results regarding student achievement and performance in regular academic settings. A special issue in *Learning and Instruction* (2013), dedicated to the application of calibration studies, illustrated several ways in which calibration can be useful to teachers' instructional decisions, but there was no empirical research on its use in classroom settings. Although it can be argued that assessments are tasks, student tasks in authentic environments such as the classroom are influenced by a multitude of factors like attitudes toward school, content area, classroom, teacher, learning, motivation, metacognition, classroom environment (physical and social), and peer relations, in addition to student anxiety and teacher characteristics. The value assigned to school tasks like assessments is different from the value assigned to tasks in an experimental study. Deliberate interventions produce results different from what can be expected in a more natural setting.

The conclusion that the setting—natural or artificial—affects the study of metacognition has at least two implications. First, our knowledge of how metacognition manifests in actual learning contexts remains limited. Second, our approach toward the instruction of metacognition will rely on designing resource-intensive interventions like formal metacognitive interventions, or creating hypermedia environments (Lajoie, 2008). A key question at this juncture inquires into ways to incorporate these extensively researched and validated techniques for measuring student metacognition into a teacher's instructional repertoire. The following paragraphs describe a handful of empirical studies that have investigated the relationship between judgments of learning and self-assessment in postsecondary classrooms as a starting point for the use of metacognitive judgments in formative assessment practices.

Boud and Falchikov (1989) examined quantitative studies on self-assessment from a variety of fields, ranging from political science to medicine, which compared college students' ratings of performance on assessments to teacher ratings. They found that, in general, the studies consulted discovered that students were accurate in their judgments; yet, there were nonetheless inconsistencies in that some studies found overrating, and some, underrating. Studies also reported a tendency for high-achieving students to underrate themselves, and low-achieving students to overrate their performance. There were no clear indicators as to whether there were differences based on field of study.

Timmers, Broekand, and Berg (2013) examined the relationship between college students' help seeking (seeking feedback and spending time on feedback) in relation their task value beliefs, expectancy beliefs, and effort on a computer-based formative assessment task. The task was labeled formative on account of its low stakes and served as an optional exercise for students to know more about their own learning. The researchers analyzed feedback behaviors

according to log data on the number of feedback pages students clicked and the time spent on each page. Students who placed a high value on the formative task expended more effort (self-reported) on the task. Timmers et al. also found a significant predictive relationship between students' expectancy beliefs (predictions on how well they performed on the task) and feedback behaviors, and that students who reported more effort showed higher feedback-seeking behaviors. The study points to the important role of student characteristics in whether they sought feedback.

In a similar study that used computer-based formative assessments to examine student self-assessment behaviors, Ibabe and Juaregizar (2010) conducted a study that examined the relationship between self-assessment and student performance among 85 undergraduate students in a data analysis class using software-based end-of-unit quizzes (multiple choice, short answer) posted after each lesson. Students took the assessment and received computer-generated feedback. The researchers also administered a questionnaire before and after the assessment. The questionnaire scaled 0-10, included items on knowledge of content. In addition, Ibabe and Juaregizar also used trace data from the software to gather information on the number of times students logged in, the duration for which they were logged in, and the number of times they took tests. They found that student judgments of knowledge significantly predicted their final grade in the course, explaining about 20% of variance in the final grade. They did not find significant differences between students who used the program and those who did not. Metacognitive variables were modestly related to performance, and students' perceptions of knowledge post-test were related to effort (measured by lecture attendance, study time, and internet use trace data). However, there was a tendency for users of self-assessment to perform better than non-users, and motivation level was positively related to frequency of self-assessment

use, indicating the potential effects of self-selection in the study. Students who participated in the study may have chosen to participate due to high motivation and positive self-beliefs.

In another study, Nietfeld, Cao, and Osborne (2005) looked at the monitoring accuracy of students throughout the semester to see if repeated metacognitive monitoring opportunities improved their monitoring accuracy. The purpose of the study was to weigh in on whether or not findings from laboratory-like studies on monitoring accuracy are generalizable to classroom-based assessment performance (McCormick, 2003). Using a small sample of 27 volunteer educational psychology students, they examined monitoring accuracy over repeated assessments for the duration of one semester. Nietfeld et al. analyzed scatter, bias indices, and correlations between accuracy, GPA and performance in individual assessments, and items within each assessment to study patterns of students' monitoring accuracy over time and across tests. They found that monitoring accuracy was fairly stable over time for both global and local (item by item) judgments of performance across the four types of assessments. They also found that global judgments on test performance were more accurate than local judgments, but that low-achieving students made weaker accuracy judgments. They found significant positive relationships between judgments and GPA, indicating that high-performing students were more accurate in their performance judgments. Although the researchers noted implications for interventions in study skills and self-regulated learning, the results may be, from a formative assessment perspective, useful for instructional practices.

Students are stable in their accuracy over time, and given that high-performing students have more accurate perceptions of their performance, students who do not perform well may have deficiencies in their monitoring skills (Schraw, 2009). Students' low accuracy on specific

test items (as compared to high-achieving students) may likewise reflect deficiencies in their monitoring skills.

Miller and Geraci (2011) conducted two studies based on experimental conditions that included providing feedback and incentives. The studies followed college students over the course of two semesters to determine whether the accuracy of exam predictions improved over time. In the first semester of the study, 130 students of an undergraduate cognitive psychology course recorded their global predictions on their performance in the form of a letter grade (A-F) before each of the four exams. Students were provided with the incentive of two extra credit points if they accurately assessed their grade. After each exam, the researchers calculated their individual accuracy scores (absolute), and also posted the average classroom grade and average exam prediction grade for the entire class to encourage students to reflect on their own performances and predictions. They found that student predictions improved slightly over time, but similar to the previous study, they found the improvement over time could not be distinguished according to whether the students were high or low performers. Students also did not improve their grades over time. Overall, however, higher performers consistently demonstrated higher calibration scores and performance scores.

In the second semester of the study, students received the same information, but were, in addition, provided explicit feedback on their accuracy in relation to their performance. They were also encouraged to be more accurate in their predictions, and were provided incentives for such accuracy. Miller and Geraci (2011) posited that students would either lower their accuracy judgments (overall, they tended to be overconfident) or raise their exam scores. They found many differences in accuracy over time for low-achieving students (especially after the second exam) compared to high performing students; students' improved accuracy was not met with

improved performance. The researchers interpreted these findings in relation to the high- and low-performing groups. High performers were consistently more accurate in their predictions, but low performers lowered their accuracy ratings to show more consistency with their performance. Although, the students may have been driven by the incentive to increase their accuracy, their ratings reflect a more realistic assessment of their performance or learning on account of the feedback they received. The next step, given these findings, is to examine how instructional support or additional feedback on study strategies may help students improve their performance. It is important to note that the researchers were instructors in the classrooms where the study was conducted. As such, the results may point to the strengths of incorporating metacognitive judgments of learning into the administration of classroom assessments.

Synthesis

The review of literature points to important student and teacher characteristics involved in the formative assessment process. Formative assessment is theorized and researched to examine how teacher and other contextual characteristics may affect its practice, and how it, in turn, impacts students and their learning (Figure 1). Sadler (1998) described teachers' expertise in the content being taught as critical to formative assessment, but also highlighted the importance of attitudinal dispositions, evaluative skills, and knowledge of standards/performance criteria. The transfer of these skills to students needs to be an explicit curriculum goal, where students have the opportunity to "consolidate and clarify their own judgments" (p. 83). However, within the formative assessment framework, little is known about the extent to which teachers are aware of and use student judgments and beliefs. Figure 2 provides a model of formative assessment that informs the purpose and design of the current study. In this study, information on student dispositions and judgments of their own learning are elicited to provide

information to teachers on the role of student characteristics and performance, and teachers' awareness and use of those characteristics. The purpose of this study is to understand formative assessment in an ESL context by examining its relationship to student and contextual characteristics using a sociocultural lens, and to explore the potential of student metacognitive judgments of learning in formative assessment practice.

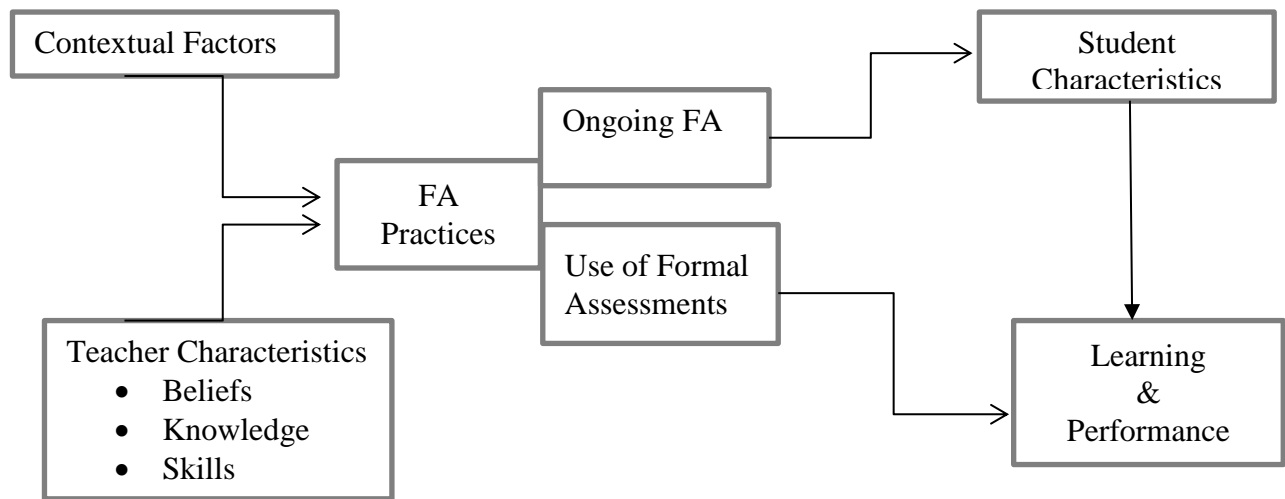


Figure 1. Model of formative assessment conceptualized in empirical research. Illustration of contextual and teacher characteristics examined in the study of formative assessment and its effect on student characteristics and student learning and performance.

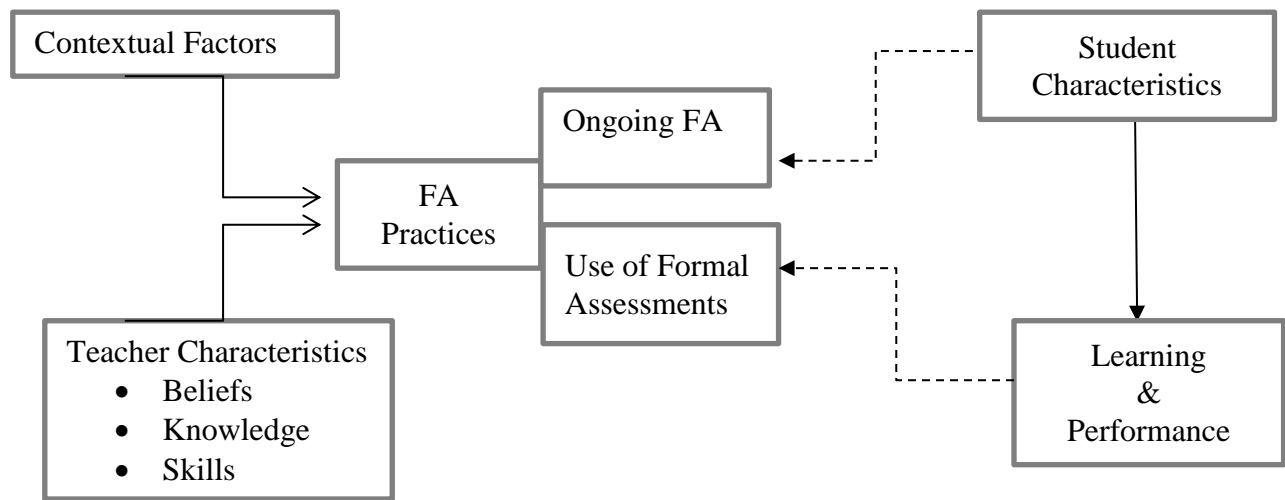


Figure 2. Model of formative assessment theory used in this study. Illustration of contextual, teacher, and student characteristics associated with the practice of formative assessment as conceptualized in the design of this study.

Chapter 3: Methodology

This chapter describes the methodology for the current study on formative assessment practices and the potential role of metacognitive judgments in formative assessment practices in ESL teaching and learning. The study design was similar to an embedded mixed methods design with concurrent qualitative and quantitative data collection (Creswell & Plano Clark, 2011). The purpose of an embedded mixed methods design is to enable in-depth investigation into a phenomenon using both quantitative and qualitative data collection methods. According to this design, one method, either quantitative or qualitative, is clearly identified as the primary strand, with the primary strand here serving to inform the philosophical assumptions of the study. The study design conforms to a common variant of the embedded design, with an emphasis on qualitative data collection methods and a smaller role of quantitative methods to investigate exploratory research questions - QUAL + quan (Creswell & Plano Clark, 2011).

One challenge inherent to the embedded mixed methods design lies in determining the timing of qualitative and quantitative data collection. Concurrent data collection is typical in embedded designs, where both components (qualitative and quantitative) may be administered at the same time. In this study, qualitative and quantitative data are collected from multiple sources, including teachers (interviews), students (surveys), and the researcher (observations). Since the qualitative and quantitative methods are implemented concurrently, the study's design components are interactive and the primary point of interface for mixing is at the design level (Creswell & Plano Clark, 2011). The procedural diagram below (Figure 3) illustrates the timing and instruments used in data collection. The following sections include the research questions guiding this study, descriptions of participants and context, measures, procedures, pilot study results, and data analysis, in addition to the potential limitations of this study.

Research Questions

Quantitative Questions

1. What is the relationship between student metacognitive judgments and student performance on unit tests?
2. What is the difference between student metacognitive beliefs at the beginning and end of the semester?

Qualitative Questions

3. What are the ongoing formative assessment practices of teachers in ESL classrooms?
4. How do teachers in ESL classrooms use classroom assessment to inform instruction?
5. How do ESL teachers use metacognitive judgment data in their formative assessment practices?

Study Overview

Sampling Procedures

The study setting was purposefully selected so as to include teachers and students in advanced level 3 written communication (WC 3) English language classes for international students. This course was selected based on several factors: 1) The course emphasizes written communication skills. Students are invested in acquiring a mastery of writing skills that will enable them to enroll and succeed in regular academic work at the university level. Apart from being an academic prerequisite, written and oral English communication skills also aid students in socially and culturally adjusting to life in the United States. The English language classroom's focus on learning, as opposed to performance, makes it conducive for the study of formative assessment. 2) The WC classes share a common curriculum, learning materials, and assessments, which allows for the comparison of different classrooms within the same course. 3)

Instructors also have the flexibility to tailor instruction to their individual classes, which allows for the examination of individual differences in the practice of formative assessment. 4) All learners were expected to be 18 years and older, which allows for minimal interference on behalf of developmental factors associated with metacognitive judgments.

Potential participants were approached after obtaining consent from the English language program director to contact teachers to discuss their willingness to participate in the study. The program director consented to the implementation of this study in fall 2014, and provided a list of instructors who would likely be teaching WC at that time. Potential participants were contacted via email to notify them of the study. In order to inform the study design, informal meetings with potential participating teachers were initiated by the researcher to gather information on the context and nature of teaching and learning. Official recruitment for the study began after approval to conduct the study from Virginia Commonwealth University's Institutional Review Board was received; recruitment was based on the fall 2014 teaching schedule for the written communication classes.

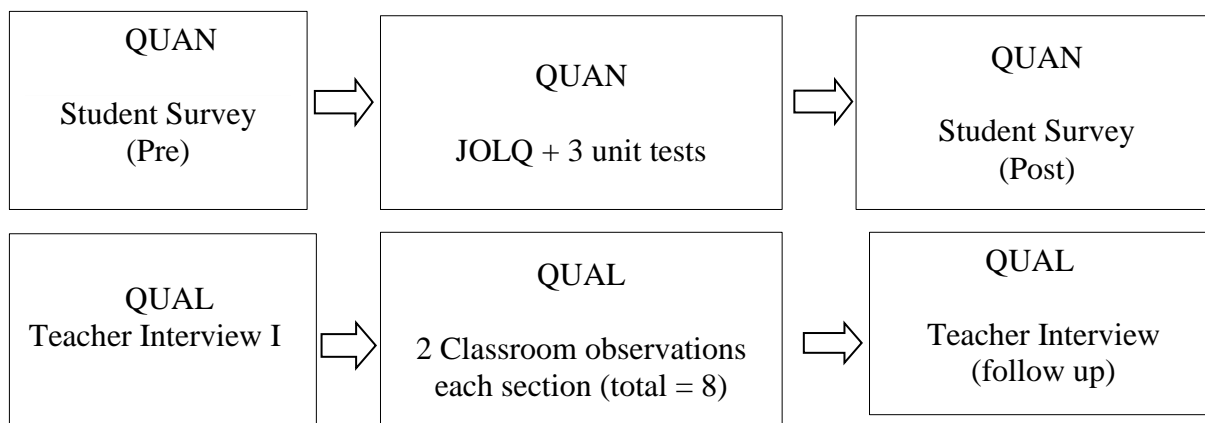


Figure 3. Mixed Methods Design Procedural Diagram. Procedural diagram illustrating the elements of the study design in relation to the timeline of implementation.

Participants

A total of three WC 3 teachers participated in the study. For the purpose of data analysis, participants were identified by their study IDs: TP01, TP02, and TP03. For the presentation of results, each teacher is identified by a pseudonym: Thomas, Daniel, and Linda; the purpose of assigning pseudonyms was to facilitate the presentation of results, and pseudonyms were chosen to reflect common American first names and the participant's gender.

Thomas, a full-time faculty member, reported 21 years of ESL teaching experience. While he has taught all domains of the English language, including speaking, listening, writing, and reading, he has taught reading and writing the most. He has a master's degree in linguistics, with a specialization in teaching English as a foreign language. During his first interview, he reported minimal assessment training, which included only one course during graduate study. He identified his skill in developing assessments as strong, and reported to have routinely prepared classroom assessments for the WC 3 courses on account of his role as lead teacher for all WC 3 sections. At the time of the study, he was teaching two sections of WC 3, both of which were included in the study.

Daniel, also a full-time faculty member, has 17 years of ESL teaching experience and has likewise taught all four domains of English language learning. He holds a master's degree in education, with a specialization in bi-lingual and multicultural education. He reported having received moderate assessment training, which included two classes in graduate school. He described also having created rubrics for his classes, serving on teacher committees, and acting as lead teacher for another level of WC classes. At the time of the study, he was teaching one WC 3 class, which was included in the study.

Linda has 20 years of experience teaching ESL students, and has taught all levels (basic, intermediate, advanced) and all domains of English language learning. She has a master's degree in teaching English to speakers of other languages. She identified her assessment training as moderate to extensive. She enjoys developing assessments, and reported her skill in developing assessments to be very strong, based on her experience designing and coordinating standardized tests. She also reported presenting her work on testing at professional conferences. She taught one WC 3 class that was included in the study. Unlike the other sections, this class had a community-based learning component, meaning that students did field work geared towards service learning. Objectives and expectations related to service learning were incorporated into the course, in addition to WC 3 learning objectives, which were identical to other classes. Even though course objectives were the same across WC 3 classes, teachers retained the flexibility to develop learning materials and assessments for the class as long as the course objectives were met. Linda's class followed a course packet and a series of informal assessments that were different from the other classes. Therefore, due to the different course materials used by the community-based WC 3 course, quantitative analysis on the metacognitive judgments of learning and performance were conducted separately for this class.

Table 3 summarizes teacher participants' educational qualifications, years of experience teaching ESL, assessment training, and other ESL classes taught.

Table 3

Description of Teacher Participants

	Thomas	Daniel	Linda
Educational Qualifications	Master's	Master's	Master's
Years of Teaching Experience in ESL	21	17	20
Self-reported Training in Assessment (None, Minimal, Moderate, Extensive)	Minimal	Moderate	Moderate-Extensive
ESL Classes Taught	Reading/Writing, Listening/Speaking	Reading/Writing, Listening/Speaking	Reading/Writing, Listening/Speaking

Note. This information was gathered during the first teacher interview

The level 3 students featured in this study were fairly advanced when it came to English language proficiency, were conversant in English, and did not require translation support to participate. Their primary goal, as level 3 students, was to strengthen written communication and/or oral communication in order to succeed in a regular undergraduate or graduate academic environment in the United States.

A total of 53 students were included for participation in the study. Thomas had 17 students in his first section (Section 1), and 13 students in the second section (Section 2). Daniel had 11 students total, and Linda had 12 students. Demographic information was collected only from students who participated in the pre- and post-surveys. Table 4 provides a summary of the demographics for students who took the pre- and post-survey. A total of 20 students responded to the pre-survey; of the 20 responses received, two were incomplete. One student contacted the researcher after receiving a link to the survey to notify that she was below 18, and would turn 18 in October. The student responded to the judgment of learning questionnaires (JOLQ)s in

October; however, she did not complete the pre- or post-survey. A total of 21 students participated in the post-survey. Since the JOLQs were administered in class on test days, the majority of students in all sections responded to at least one JOLQ, $N = 51$.

As indicated by the pre-survey, 63% (13/21) of students who participated in the study were between the ages of 18 and 26; 4 students were between 26 and 35; and 3 students were between 36 and 45. At the time of the post-survey, 13 respondents were between the ages of 18 and 25, and 5 (21%) were between 26 and 35; 3 students were between 36 and 45. In addition to age, participants also reported student status in the post-survey. A majority of respondents (67%) were full-time English language program (ELP) students; four (16%) were full-time undergraduate students, and a smaller percentage, two students, (8%) were graduate students, and one student chose “Other,” but did not specify his/her status. The respondents represented 10 countries: China, Columbia, Iraq, Japan, Kazakhstan, Romania, South Korea, Saudi Arabia, Spain, and Vietnam. However, the ELP student body was predominantly from Saudi Arabia. In fact, about 32% of study participants disclosed that they were from Saudi Arabia. In the pre-survey, 50% of students (10/20) reported that they never or rarely talked to friends and family in English, whereas about 30% said they did sometimes. In the post-survey, about 40% of students reported that they talked to friends and family in English “sometimes,” but about 50% responded with “rarely” or “never.”

Table 4

Description of Student Participants

Demographic	Pre-Survey Count (%)	Post-Survey Count (%)
Age		
Below 18	0	0
18- 25	13 (63)	13(62)
26-35	4 (21)	5(24)
36-45	3 (16)	3(14)
Above 45	0	0
Gender		
Male	11(55)	10(48)
Female	8 (45)	11(52)
Student Status		
Full-time ELP student	-	14(67)
Full-time undergraduate student	-	4(19)
Full-time graduate student	-	2(10)
Other (not specified)	-	1(5)
Country of Origin		
China	1(4)	3(12)
Columbia	1(4)	1(4)
Iraq	2(8)	2(8)
Japan	2(8)	2(8)
Kazakhstan	0	1(4)
S. Korea	2(8)	2(8)
Romania	0	1(4)
Saudi Arabia	8(32)	8(32)
Vietnam	0	1(4)
Spain	1(4)	0
How often do you talk to your friends/family in English?		
Never	1(5)	4(17)
Rarely	9(47)	6(26)
Sometimes	6(32)	9(39)
Often	1(5)	4(17)
Always	2(11)	0

Context

The level 3 written communication (WC 3) class is a required class for all students enrolled in the ELP. The program is led by a director, an associate director, and an assistant director, and is comprised of approximately 40 full-time and part-time teaching faculty members.

The ELP is an accredited member of the American Association of Intensive English Program (www.englishusa.org) and the Colonial Academic Alliance (www.colonialacademicalliance.org), organizations that promote educational quality for intensive English language programs and institutions, respectively. The program offers three levels of English language classes—beginner (level 1), intermediate (level 2), and advanced (level 3)—in four core areas of language learning, including reading, writing, listening, and speaking. The program also provides academic preparatory courses (e.g., psychology, advanced Graduate Record Examination (GRE) preparation, engineering, etc.) within an EFL setting. Classes generally last 85 minutes each, and meet five days a week. The average class size for each written communication section is about 12 students, and classes are typically interactive, providing plenty of opportunity for students to engage in and practice the English language. The course materials used across the different classes include *Q Skills for Success*, a series of books and resources in all core areas of English learning (with the exception of the WC 3 community-based class, as noted above in participant descriptions). Of the four participating classes, three were scheduled 9:00am to 10:25am, and one class was scheduled 11:00am to 12:25pm. All classes met five days a week in the same classroom, with the exception Linda’s class, which engaged in service learning activities outside of class on Wednesdays, in addition to a class meeting that same day. To compensate for the extra hours dedicated to Wednesdays, the community-based WC 3 class did not meet on Fridays.

Assessments

While the three WC 3 sections included in this study used common midterm and final assessments, a small percentage of the exams administered in each section remained variable. The WC 3 courses adopted an 80-20 approach, wherein the lead teacher for the course collaborated with all WC 3 teachers to develop a midterm and a final assessment where 80% of items were common across sections and instructors, and 20% of exam material could be tailored to each instructor's class. Together, the midterm and final exam represented 20% of the final grade in the course. Four end-of-unit classroom assessments were developed and administered using processes and procedures similar to those employed for the midterm and final exams. Unit assessments were administered to monitor student progress after each content unit, and accounted for 20% of students' course grade. Each assessment, including end-of-unit tests, midterms, and final exams, was administered over the course of two days. The first day included an essay component, where students responded to a writing prompt. The second day of testing included a reading comprehension and grammar/vocabulary skills component, where students responded to questions based on a reading passage. Instructors were granted the flexibility to determine the instructional activities and graded classroom assessments for their section. A significant portion of the course grade (up to 50%) was allocated for portfolios, which include both in-class and take-home writing assignments. Class participation and homework accounted for 10% of the student's final grade.

Measures

Teacher interviews. Two semi-structured interviews were conducted with each participating teacher. The first interview was conducted in September/early October 2014. The follow-up interview was conducted in late November/early December. The researcher developed

two protocols for the interviews (see Appendices A and B). Descriptions of the protocols are provided below.

ESL teacher assessment practices interview (TAPI - I). This protocol was developed for use in a semi-structured interview. The protocol included 17 main questions divided into seven sections. The first section, the background section, assumed the form of a questionnaire to collect descriptive information on teaching experience, typical instructional activities in WC 3, and teachers' self-ratings of their skill in developing and using assessments; teachers' rated their skill level on a four-point Likert-type scale (very weak, weak, strong, very strong). The second section included questions on participants' grading and assessment practices in general, as well as specific plans for WC 3. This section also measured the extent to which program policy influenced the abovementioned practices by using a four-point Likert-type scale (never/rarely, sometimes, often, always), as well as follow-up questions to elaborate on the participant's chosen response. The third section of the protocol included two questions on participants' beliefs about the characteristics of students who earn A grades, and those who earn F grades, and the typical proportion of students earning As and Fs in the class. The fourth section inquired about the ongoing formative assessment practices of participants and included three subsections, viz., learning goals, eliciting evidence of learning, and making instructional adjustments (Heritage, 2009). This section included questions related to the frequency of instructional adjustments; response options were formulated according to a four-point Likert-type scale (never/rarely, sometimes, often, always), with follow-up questions for respondents to elaborate on their chosen response. The fifth section again used a four-point Likert-type scale (never/rarely, sometimes, often, always), and featured follow-up questions for participants to elaborate on their responses. This section included questions on participants' use of formal assessments to inform instructional

practices, their feedback practices on formal assessments, and challenges associated with using assessments for instructional adjustments. The sixth section included questions on participants' perceptions regarding the nature and role of student characteristics in teaching and learning, including motivation, metacognition, and student use of feedback. The final section is a concluding question, which asks participants to share anything related to assessment practices that may have been missed in the interview protocol. The researcher tested the protocol by conducting an informal interview with a WC 3 teacher who was not a participant in the main study. The teacher provided feedback on redundancies, difficulty responding to any question, and the appropriateness of the questions for WC 3 ESL teachers in the study context. The protocol described above reflects revisions made based on the pilot participant teacher's feedback and a discussion with the dissertation chair.

ESL teacher assessment practices interview TAPI - II. The second interview protocol was likewise developed by the researcher. This protocol included 21 questions, several of which followed up on questions issued in the first protocol as a means of determining whether there were changes to participant responses in the first interview and to document any changes that occurred during the semester. The protocol was divided into seven sections. The first section included questions on typical activities used in the WC 3 classroom, and the average preparation time for required classes; this section also revisits assessment development and use, measuring responses according to a four-point Likert scale, with response options ranging from very weak to very strong. In order to follow up on any changes to WC 3 for the fall semester, the second section summarized participant responses in the first interview regarding their grading and assessment practices. The third section followed up on any changes regarding participants' descriptions of those WC 3 students who receive A and F grades. In addition, in this section,

participants were asked to describe the learners in WC 3 based on the information they received from the student survey and the judgments of learning data. The fourth section is identical to the fourth section of the first interview, with the exception of language-related changes made to questions about participants' formative assessment practices for WC 3, specifically. In addition, here, teachers were also asked to describe whether and to what extent the judgments of learning data informed their formative assessment practices. The fifth section was identical to the fifth section of TAPI-I, described above. Two additional questions, however, were included in this section of the second protocol; these questions asked teachers to describe any changes in the nature, frequency, or type of feedback provided to students based on access to student survey data and judgments of learning data. Similar to the fourth and fifth sections, the sixth section of was identical to the sixth section of the TAPI – I, with one exception. The sixth section of the TAPI-II added questions related to the consistency of survey and judgments of learning data in relation to teachers' perceptions of student characteristics, such as self-beliefs and metacognition. The final section represented an opportunity for participants to share anything that may have been left out by the interview protocol.

ESL Student Metacognitive Beliefs Survey. Student participants took two surveys at the beginning and end of the semester (see Appendix C). The pre- and post-surveys were identical; however, the pre-survey included the consent form detailing the purpose of the study and the conditions for participation, including a response area where students could “agree” or “disagree” to participate in the study.

The surveys contained 25 questions, including five demographic questions on age, gender, country of origin, student status at the university, and students' social use of the English language. Survey questions also asked about the perceived difficulty of the four domains of

English language learning, listening, speaking, reading, and writing, and included an eight-item scale on metacognitive beliefs related to writing (Cronbach's $\alpha = .90$), a five-item scale on influences on writing (Cronbach's $\alpha = .57$), and a two-item scale on the application of coursework in class (Cronbach's $\alpha = .70$).

Questions were developed by the researcher with the purpose of conducting two pilot administrations. The first pilot administration was an informal administration of the survey in spring 2014, and included a convenience sample of participants known to the researcher to be from diverse educational and demographic backgrounds. The purpose of this administration was to identify redundancies and reduce survey items for the second pilot study. The survey was comprised of several items and scales related to student conceptions of learning, assessment, and metacognitive beliefs about writing. Since the obtained sample size was less than 100, only metacognitive beliefs related items were retained for the main study. Decisions regarding the items in the scale were based on exploratory factor analyses results. Results indicated a sound factor structure for the scales that were eventually included in the main study after testing the final scales in a WC 3 classroom during an informal administration. The second pilot administration was conducted in July 2014, and included a convenience sample of adult learners similar to the actual sample of the study in terms of demographic and educational background. The participants in the second pilot were students in the same program as those featured in the main study, but were not included in the main study sample. The results of the pilot were used for the single purpose of data analysis to inform instrument development.

Observations. The researcher conducted two classroom observations for each teacher/class in order to examine ongoing formative assessment interactions in the participating classrooms. A basic observation protocol was developed to record ongoing activities and

descriptions of the classroom environment as well as describe the nature of formative assessment interactions (see Appendix D). In summer 2014, the protocol was tested in one pilot classroom observation in the same WC 3 classroom where the initial observation was performed. The purpose of the pilot observation was to determine the ability of the observation guide to record specific formative assessment interactions in the classroom. Based on the pilot administration, it was determined that observation sessions in the main study would be audio recorded to facilitate detailed note taking after the observation; additionally, the guide would serve as a tool to summarize key formative assessment interactions in order to facilitate building the narrative of results. Classroom observations were included for the following purposes: 1) to support or enhance interview data on participant formative assessment practices; 2) to provide insight into the nature of classroom interactions as they relate to the language of formative assessment used in the literature; 3) to understand the differences and similarities characterizing participating teachers' instructional practices.

The observation guide included three sections. This first section was dedicated to the setting. It included a place to record the number of students, a description of the physical setting, and observations of students' behavioral engagement using field notes. The second section included a table to record summaries of activities taking place in the classroom, as well as the duration of each activity documented, and field notes to accommodate additional details. The third and final section of the guide was concerned with formative assessment interactions. This section was based on McMillan et al.'s (2013) framework for analyzing formative assessment practices in empirical studies of formative assessment, and Panova and Lyster's (2002) work, which uses Spada & Frohlich's (1995) well-known observation protocol for English language communication and feedback in the ESL classroom. This section included a checklist to record

the origin of interactions (teacher/student initiated), nature, type, specificity, timing of feedback interactions, and student response to feedback.

Judgments of learning questionnaire (JOLQ). The JOLQ administered to student participants included four statements, which were developed by the researcher based on typically used metacognitive judgments and monitoring measures (e.g., Hattie, 2013; Huff & Nietfeld, 2011). For each question, the statements included response options on a scale of 1 to 10 (see Appendix E). Examples of the questions featured in the JOLQ include: How difficult is the writing test given to you today? (not at all difficult–very difficult); How well did you prepare for today’s test?(did not study at all–studied very well); How well do you know the material from Unit X? (not well at all–very well); How confident are you that you will do well on this test? (not at all confident–very confident). An optional open-ended item was included for students to share any thoughts or feelings about their preparation or perceptions of the test. The JOLQ was pilot-tested with one WC 3 class from the same program in July 2014. Data at the class level (no identifying information for individual students was collected) were shared with the classroom teacher to get feedback on the usefulness of the information garnered from the JOLQs in relation to student performance on the assessment. Since summer classes are short and hectic, the teacher was not able incorporate any changes based on the summary data. However, the teacher provided feedback to the researcher on the language used in the four items relative to students’ English language proficiency and approximate time students took to complete the questionnaire. The teacher suggested a minor change in language of the JOL item on knowledge- he suggested that the language be clarified from “how well do you know *about* the material taught in Unit X?” to “how well do you know the material taught in Unit X?” Students approximately took two minutes to complete each questionnaire. For Linda’s class, the wording of the questionnaire was

modified since her students did not take end-of-unit tests; unit numbers were replaced with “in today’s in-class writing” or “reading comprehension quiz test.” For three unit tests, the students took the JOLQ prior to starting the test, but after they received the test forms. In total, the JOLQs were administered six times in each class (three times in WC 3), including two administrations (one for the writing component, and one for skills/reading component) for each unit test taken over the two-day period.

Student performance data. Teachers provided student scores on each of the three assessments as well as their final course grade. These measures were used as indicators of performance and achievement in the data analysis procedures to address the quantitative research questions.

Procedures

Teacher participants were contacted based on a shortlist provided by the program director. Three teachers expressed interest in participating and met the researcher in person to go over the nature and purpose of the study, during which time the researcher described the procedures involved and their rights as participants. Participants signed the consent forms during the first teacher interview. All interviews were conducted in person in the participant’s office or in a small conference room at the university library. At the first interview, the researcher obtained a tentative schedule for the three unit tests. The researcher also identified a convenient class session to introduce herself to the students, explain the purpose of the study, and request their participation. All three first-round interviews were audio-recorded after obtaining permission from participants. Interview tapes were sent to a professional transcriptionist to convert the audio, verbatim, into written transcripts.

The researcher conducted her observations after consulting with each teacher to identify a day that would be a typical class day. Observations were conducted in early October and late November. The researcher audio-recorded class sessions in order to take detailed field notes after the observation session. After each observation, the researcher transcribed important interactions in a Word document, noting comments and other relevant information about the classroom environment.

The researcher visited all four classes to talk to student participants, inform them about the study, provide them with an opportunity to ask questions or raise concerns regarding their participation, and also ensure that they understood and were comfortable with all aspects of the study. The researcher told them they could expect a link to an online survey from the researcher's VCU email ID in the next 24-48 hours. The researcher sent a personalized email to each student participant with a link to the pre-survey using the RedCAP survey platform. The online pre-survey was administered over a three-week period in early October 2014; those who did not respond were sent reminders once every three to four days for the first week.

The JOLQs were conducted via paper-and-pencil forms that were handed to the teachers one day before each unit assessment. The teacher administered the JOLQ at the beginning of the test, after students received the test forms, but before they started the test. The students took approximately two minutes to complete the form. Upon completion of the test, the teachers collected the forms and handed them to the researcher. The teachers provided student grades on the unit tests 3 to 14 days after the test. The researcher provided teachers with summary reports of the JOLQ at the class level after receiving student grades on the unit assessments. The summary report included charts that contained student ratings of each of the four questionnaire items; charts were created using Microsoft Excel (a sample summary report is provided in

Appendix F). The summary also included correlations between student ratings on each JOLQ and their score on the test in order to understand the accuracy of student JOL in relation to test performance. Teachers were encouraged to reflect and use the information on metacognitive judgments as they saw fit, which included using it to inform their feedback to individual students, to inform instructional activities, or to facilitate class discussion.

The final interviews and student surveys were conducted during the last two weeks of the semester, in late November and early December. The 20-25-minute student survey was administered online using RedCAP survey software; the survey was open until the last week of December. The follow-up interview with teachers was conducted in participants' offices or a group study room in the university's library in early December 2014.

Data Analysis Procedures

Quantitative data procedures. Quantitative data collection included two administrations of the survey and six administrations of the JOLQ (four administrations for Linda's class). The pre- and post-surveys administered online using RedCAP survey software included a link unique to each participant as a way for the researcher to identify and track each participant over multiple data points, viz., post-surveys and JOLQs. Once the survey was closed for participation, complete and partial data were downloaded in Microsoft Excel format for cleaning. Once the researcher ensured that data had transferred accurately, it was imported into SPSS statistical analysis software. All statistical analyses to address the research questions were conducted on SPSS software.

After obtaining the JOLQ results, the researcher manually entered the data points into a Microsoft Excel file. After spot-checking at least 30% of entries for accuracy, the researcher prepared charts and a narrative interpreting of the results for each teacher in a Word document.

At the end of the study, the researcher merged responses from all JOLQs for analysis related to the study's research questions.

The following paragraphs provide a description of quantitative data analysis procedures for this study's quantitative research questions.

What is the relationship between student metacognitive judgments and their performance on unit tests? This question was addressed by examining bivariate Pearson correlations between student responses on the JOLQ and their corresponding assessment scores. Since JOL regarding difficulty, performance, and preparation for the test are provided at the global level, correlation coefficients represent calibration scores and are appropriate for interpreting the relationship between JOL and performance scores. Calibration scores can be positive or negative, between -1 and +1, where high positive/negative correlations indicate higher accuracy, and low correlations indicate inaccuracy (Schraw, 2009). For example, JOL regarding difficulty of the assessment can be expected to correlate negatively with performance scores if learners are accurate in their metacognitive awareness and monitoring.

What is the difference between student metacognitive beliefs at the beginning and end of the semester? This question was addressed by analyzing pre- and post-survey data according to the metacognitive beliefs scale using paired *t*-tests to examine changes in student conceptions at the beginning and end of the semester.

Qualitative data analysis procedures. The procedures for analyzing interview data and observation data followed those outlined by Bogdan and Biklen (2007). The researcher used ATLAS.ti qualitative data analysis software to process, code, and interpret findings from interviews and observations. Using audio tapes of classroom observations, observation field

notes were first composed in narrative form in a Word document, and then later uploaded to ATLAS.ti for analysis.

Each interview was coded using an iterative process, with coding procedures informed by the work of Bogdan and Biklen (2007) and Lewins and Silver (2007). Coding categories included context-related codes, participant perspectives, strategy codes related to formative assessment, process codes for comparing multiple interviews with the same participant, and process codes in relation to formative assessment theories of eliciting evidence, making sense of the evidence, and taking action. Much like the coding used for interviews, the researcher used similar coding for observation data. Codes were analyzed at a second level with the intention of reducing the number of codes, and eliminating or merging redundant codes. Themes were identified based on frequency and strength of responses for codes, but also according to the significance of the code in the participant's response. The following paragraphs outline the data analysis procedure for qualitative data by research question.

What are the ongoing formative assessment practices of teachers in ESL classrooms?

The above question was primarily addressed by interview and observation data. In responding to this question, coding and analysis procedures categorized emerging themes into categories or code families that represented an overarching topic. Using both sources of information (interview data and observation data) together, the researcher was able to describe similarities, differences, and peculiarities in participants' formative assessment practices.

How do teachers in ESL classrooms use classroom assessment to inform instruction?

Interviews with participating teachers constituted the primary source of information for addressing this question. In fact, the fifth section of the interview protocol featured questions related to teachers' use of formal assessment data. In Chapter 4, themes related to teachers' use

of formal assessment will be described to address this question; additionally, at this time, the discussion will also attend to relevant themes that emerged from the previous research question, ‘What are the ongoing formative assessment practices of teachers in ESL classrooms?’

How do ESL teachers use metacognitive judgment data in their formative assessment practices? The second interview conducted with teachers represented the primary source of information for addressing this question. More specifically, questions related to teachers’ use and perceptions of the JOLQ summary data were used address this question.

Validation Procedures

Quantitative data. Both the student survey and the JOLQ were pilot-tested in summer 2014 in order to determine the reliability and validity of the scores. The data were analyzed using a principal component analysis or exploratory factor analysis to support construct validity. The findings from this analysis then informed subsequent revisions to survey items. Content-related validity is established by aligning survey items to validated measures. The survey was reviewed by a WC 3 teacher not participating in the study to ensure the appropriateness of the language used in the survey and JOLQ.

Qualitative data. The interview protocols and observation guide were developed to closely align questions with the conceptual framework informing the present study. In order to establish trustworthiness, one ESL teacher participated in an interview with the explicit purpose of providing feedback on the questions included in the protocol. One pilot observation was conducted with the same teacher’s class to inform the use of the observation guide. In addition to instrument development, validation procedures for data analysis included triangulation with observation data, and assessing the inter-rater agreement of initial codes generated during data analysis. An independent coder well versed in qualitative education research methods and higher

education teaching and learning coded one interview and one observation transcript. This was followed by a discussion with the researcher about similarities and discrepancies in coding. Details related to inter-rater agreement are provided in Chapter 4, page 116. The following section elaborates on the steps taken by the researcher to bolster the trustworthiness of the study.

The Researcher's Role. A significant portion of the proposed study involved fieldwork that included participant observation and in-depth interviews. The success of these techniques in gathering relevant, meaningful information depends on the interviewer's competency and knowledge. Bogdan & Biklen (2007) provided strategies to help researchers gather meaningful and high quality data from interviews and observations. They address the role of the researcher's characteristics and researcher's own biases that may influence the nature of the information collected. They suggest researchers to be discreet and try to blend into an environment they are observing to reduce the possibility of altering behaviors of participants in a way that is not typical in that setting. All participants knew the objective of the study and the purpose of classroom observations. The researcher also made it clear that she wants to observe typical classroom interactions, and was not observing for the purpose of evaluating learning and instruction. By positioning herself as someone who is observing to learn more about the environment, the researcher minimized the change in behaviors that may occur as a result of the researcher's presence in the classroom. The pilot observation of one classroom similar to the study setting helped the researcher become familiar with elements of the classroom culture that can be expected, and the revision of the observation protocol to make it more tailored to gather the most relevant information. Because of the fast pace with which classroom interactions occurred, the researcher decided to audio record classroom interactions for the purpose of aiding

note taking. This helped the researcher focus on key interactions to observe and record as opposed to taking down verbatim notes of interactions.

With regard to the researcher's own feelings and biases, Bogdan & Biklen (2007) recommend that researchers record their feelings before and after observations and interviews as a design strategy to reduce bias. The researcher used journaling as a technique to record thoughts and feelings prior to and immediately after interviews and observations. Journal notes were not used in the data analysis, rather, they served as important indicators or reminders of events or situations that may need to be considered while analyzing data and/or help prevent researchers' own feelings from seeping into analyzing the data.

The researcher's pre-existing relationship with all teacher participants of the study can be characterized as formal and courteous, where conversations have typically been on the study topic. Therefore, the researcher anticipated participants to be comfortable with the interview questions. Also, during the pilot interview with one teacher, the researcher discussed whether questions seem appropriate and answerable. The participant's feedback was incorporated in the revision of the interview protocol. Going over the protocol with the pilot participant also helped the researcher get familiar with the flow of the questions.

Institutional Review Board

Prior to data collection, this study proposal was submitted to the Virginia Commonwealth University Institutional Review Board (VCU IRB) for an expedited review. Expedited review is recommended for human subject studies involving procedures described by federal regulations as posing minimal risk to participants. The study met the requirements for an expedited review and was approved in September 2014 (Approval #HM20002382). In an effort to increase student participation in the study, an amendment to the initial application was submitted to the IRB so as

to include incentives for students. The amendment also requested that the researcher obtain student consent in class for students who did not take the pre-survey online. The incentives included one of the following the university's official merchandise, each priced below five dollars – a magnet, keychain, or a tote with the university logo. In addition, one student would be selected from the list of participants to win a twenty five dollar Amazon.com gift card in a raffle at the end of the study. The amendment was approved by the IRB in September 2014. A continuation of study request was submitted to the IRB in July 2015, one year after initial approval. This was approved by the IRB in August 2015.

Ethical Considerations

The nature of data collection procedures warrants special consideration to protect the identity and interests of study participants. Procedures put in place to protect participant identity include the following: 1) Only the researcher had access to information that could identify participants. 2) As soon as data from different sources were consolidated, the data were de-identified and a study ID was assigned to each participant prior to conducting any data analysis. All protocols included participant IDs, and the file connecting participants' identities to their study IDs was kept in a secure location accessible only to the researcher. 3) Student consent and instruction forms stated that participation or non-participation would not affect students' grades in any way. These statements were emphasized by using bold typeface or underlining. Furthermore, when the researcher met the students in class to explain the purpose of the study, she provided students with a copy of the consent form and thoroughly discussed its contents. Providing students with a copy of the informed consent form well in advance of their participation in the study ensured that students had time to review the conditions of their participation. 4) Since all students were English language learners, it was important to meet the

students prior to the study to discuss their participation and provide time to clarify their questions. 5) All attempts were made in the study design to minimize any disruption to instructional and assessment routines. 6) Informed consent was obtained from teachers prior to the first teacher interview and from students during the first online survey and prior to the administration of the first JOLQ in class.

Incentives were handed out to students during the last week of the semester in December 2014. The researcher notified students to collect the incentives at a common university location during specific times over two days.

Study Delimitations

The purpose of this study is to meaningfully contribute to the knowledge of formative assessment practices, as well as knowledge of the role of metacognition in supporting formative assessment. This study's findings are delimited to the population and the context in which the study occurred. The study was designed as a case study, which does not aim to generalize to all educational settings at the higher education or primary/secondary level. As such, the data will not represent instructional practices or student characteristics in settings outside ESL classrooms in the United States. Furthermore, international student participants may possess a cultural background that is different from typical university students and English language learners in other universities.

Chapter 4: Results

The current chapter presents this study's results in the order of its research questions, beginning with quantitative questions, followed by qualitative questions. The quantitative results for the first research question are divided into two sections. The first section groups results from student responses for Thomas and Daniel since both teacher participants used common unit assessments in their classes. The second section includes results from student responses for the third teacher participant, Linda, who administered classroom-based assessments that were different from Thomas's and Daniel's sections. For each section, results are presented in the order of the unit tests: Unit 6, Unit 7, Unit 9; for Linda's students: Midterm, Writing Test 2, and Writing Test 3.

Both the quantitative and qualitative research questions (RQ) for this study are presented below.

Quantitative Questions:

RQ1. What is the relationship between student metacognitive judgment and their performance on unit tests?

RQ2. What is the difference between student metacognitive beliefs at the beginning and end of the semester?

Qualitative Questions:

RQ3. What are the ongoing formative assessment practices of teachers in ESL classrooms?

RQ4. How do teachers in ESL classrooms use classroom assessment to inform instruction?

RQ5. How do ESL teachers use data on metacognitive judgments of learning data in their formative assessment practices?

Quantitative Results

What is the Relationship between Metacognitive Judgment and Test Performance for Thomas and Daniel's Students?

To address this research question, bivariate correlations between each JOL and its corresponding writing and skills test score were examined. The researcher first examined the association between each of the four JOL items. A high positive correlation was expected between preparation, knowledge, and confidence ratings, and a negative correlation was expected between difficulty ratings and the other three JOL. Correlations indicated that for each test, student ratings on each JOL were significantly and positively correlated to each other in every writing and skills test component for preparation, knowledge, and confidence. Every correlation was significant; Pearson r ranged from .38 to .90. The relationship between the three JOLs and difficulty ratings, however, were harder to interpret; for all three writing test ratings, difficulty ratings were not related to the other JOLs. For reading/skills tests, difficulty ratings demonstrated a significant positive correlation to all other corresponding JOLs in the second unit test, the corresponding knowledge rating of the first and second unit test.

Prior to analyzing the correlations, scatterplots for each relationship were examined to check for curvilinear relationships and the presence of outliers. No significant concerns were found, except for JOL and skills test scores for units 6 and 7. In these instances, students marked very low or very high JOL ratings, yet exhibited very low (below 50) skills test scores. For example, a student marked the Unit 6 skills test as very difficult, rating it a 10. Further, this student's score on the skills test was commensurate to the perceived difficulty of test, as he/she

received a 43. At the same time, however, three other students in the group earned similar test scores, yet their difficulty ratings were in the low to moderately difficult range, consistent with the rest of the group.

In another instance, for the Unit 7 skills test, a student who earned a very high test score (above 90) indicated little preparation and low confidence in his/her JOL ratings. Another student in the same group rated very low knowledge of the material and low confidence (rating of 2 on a scale of 1-10), and scored 42 on the corresponding skills test. Since the data points were valid and relevant indicators of accurate judgments of learning, outliers were retained in correlational analysis.

The sample size available for the correlational analysis was $n = 41$. Missing data were handled using the default pairwise deletion option on SPSS for correlations. As a result, the sample size for individual correlations ranged between $n = 28$ and $n = 36$. Table 5 provides the sample sizes for each instrument administered to student participants of the study.

Table 5

Sample Size for student surveys and JOLQ

Instrument	Sample Size*
Judgments of Learning Questionnaire (JOLQ)	51
Thomas and Daniel's sections	41
Linda's section	10
ESL Student Metacognitive Beliefs Survey (Pre)	21
ESL Student Metacognitive Beliefs Survey (Post)	25
ESL Student Metacognitive Beliefs Survey (Pre and Post)	16

Note. Sample size includes incomplete responses that were excluded from analysis

Table 6 illustrates the correlations between JOL and test scores for the three end-of-unit assessments. The rows represent the four JOL administered to students at the beginning of each of the two components for each unit test, writing and reading. The columns represent student scores on the three unit tests that were included in the study; each test featured in the column is

subdivided into a writing score and a reading/skills score. While the primary aim was to look at the relationship between the JOL provided for the writing test in relation to the writing score, and the JOL provided for the skills test in relation to the skills score, there were significant correlations between JOL for writing and skills test scores that may be relevant to the research question at hand. These correlations are noted in the description of results below.

Unit 6

The tests scores for the writing and skills components were moderately and significantly correlated, $r = .47, p < .01$. Student scores for the writing test showed considerable variability: the lowest score was 0 ($n = 2$); the highest score was 92; and the mean score was 66.15. The lowest score for the reading/skills test, on the other hand, was 0; the highest score was 100 ($n = 1$); and the mean score was 67.11.

For the writing component of Unit 6, students' JOL for their preparation (how well did you prepare for today's test?) did not show any correlation to the writing score with $r = -.00$, and with $r = .08$ for the reading score. Knowledge JOL for writing was significantly and positively correlated with writing and reading scores, $r = .41, p < .05$ and $r = .35, p < .05$, respectively. Students' ratings on test difficulty demonstrated low correlations with the writing score ($r = .18$) and reading score ($r = .13$). Similarly, low correlations were also noted between students' confidence ratings and their writing and reading scores: $r = .03$ and $r = .32$, respectively. For the skills component JOLs, preparation on the skills test positively correlated with writing test scores, $r = .38, p < .05$, but not the skills test, $r = .22$. Similarly, knowledge JOL, or how well they knew the material taught in Unit 6, was significantly correlated with the writing score, $r = .45, p < .05$. Difficulty ratings on the skills test displayed a similar significant positive correlation with the writing test score ($r = .39, p < .05$), and a low correlation with the skills score, $r = .11$.

Although moderate, correlations between confidence ratings on the skills test and writing and reading scores were not significant, $r = .34$ and $.24$, respectively.

Unit 7

The writing and skills test scores indicated variability. For the writing test, the lowest score was a 33, and the highest score a 96, with the mean writing score increasing to 74.57. For the skills test, the lowest score was 34; the highest score was 100 ($n = 1$); and the mean score was 74.5.

The test scores for the writing and skills components were moderately and significantly correlated, $r = .40$, $p < .05$. For the writing component JOLs, preparation was not related to the writing or reading score ($r = -.08$ and $.07$, respectively); knowledge ratings were significantly correlated with the skills score ($r = .38$), but not the writing score ($r = .09$); difficulty ratings showed no relationship with the scores ($r = -.04$ for writing; $r = .02$ for skills); and confidence ratings showed moderate correlation that did not reach significance for the skills test ($.34$), and exhibited very low correlation with the writing test score ($r = .03$). For the skills component JOLs, preparation was not related to writing or skills test scores, $r = -.08$, $r = .25$, respectively; knowledge was related to the writing test score ($r = .39$, $p < .05$), but there was also a correlation with the skills test score ($r = .05$); difficulty was not related to the writing or skills score ($r = .24$ and $.29$, respectively); finally, there was no significant relationship between confidence ratings and writing test scores ($r = .25$), but there was a significant moderate correlation between confidence JOL for the skills test and the corresponding skills scores ($r = .45$, $p < .05$)

Unit 9

For the Unit 9 assessments, the group's average test score further increased to 80.71 for the writing component, with the lowest score being a 58, and the highest score a 98. For the

skills component, the average score for the group dropped to 70.87. The lowest score was 12.5 ($n = 1$), and the highest score was 100 ($n = 1$). The skills and writing component test scores were significantly correlated, $r = .66, p < .01$.

For the writing component JOLs, there was no correlation between any of the four JOLs and test scores. For the reading/skills component, there was no correlation between preparation, difficulty, or confidence ratings and test scores; the knowledge rating was significantly and moderately related to the skills test ($r = .40, p < .05$), and revealed a low correlation with the writing test ($r = .20$).

Table 6

Bivariate Correlations between Judgments of Learning and Unit Test Scores for Thomas and Daniel's Students

JOL		Test 1		Test 2		Test 3	
		Scores(Unit 6)		Scores(Unit 7)		Scores(Unit 9)	
Writing Component		Writing	Skills	Writing	Skills	Writing	Skills
	Preparation	-.004	.08	-.08	.07	.19	.11
	Knowledge	.41*	.35*	.09	.38*	.31	.27
	Difficulty	.18	.13	.13	-.04	.02	.10
	Confidence	.03	.32	.03	.34	.33	.20
Reading/Skills Component	Preparation	.38*	.22	-.08	.25	.17	.30
	Knowledge	.45**	.34	.05	.39*	.31	.40*
	Difficulty	.39*	.11	.24	.29	.10	.03
	Confidence	.34	.24	.25	.45*	.20	.29

Note: * significant at .05 level; ** significant at .01 level (2-tailed)

What is the Relationship between Metacognitive Judgments and Test Performance for Linda's Students?

Because Linda's class was a written communication class with a community service learning component, the class did not follow the same set of formal unit assessments as the other WC 3 sections. As a result, the teacher administered the JOL questionnaire for similar assessments conducted in lieu of unit assessments, which included one midterm test with a reading and writing component, and two in-class writing exercises. The class strength was 12 students, of which data were available for 7 students for the writing component of the first test, with 10 students having taken the reading/skills portion of the test, and 8 students having taken writing tests 2 and 3. The JOL questionnaire was modified in consultation with the class teacher to reflect the language used by the class to reference the assessment; "test" or "unit test" was replaced with "in-class writing"; the reading/skills test was renamed "Reading Comprehension Questions Quiz"; the knowledge JOL question was modified to read, "skills to be able to do well," in place of "material that was taught in Unit X."

The four JOL questions for the writing test were:

1. How well did you prepare for today's in-class writing? (did not prepare at all-prepared very well)
2. How well do you know the skills to be able to do well on the in-class writing? (not well at all-very well)
3. How difficult is today's in-class writing? (not difficult at all-very difficult)
4. How confident are you that you will do well on today's in-class writing? (not at all confident-very confident)

The four JOL questions for the reading comprehension quiz were:

1. How well did you prepare for today's test?
2. How well do you know the answer to the questions for today's test?
3. How difficult is the test given to you today?
4. How confident are you that you will do well on this test?

Although several moderate positive and negative correlations can be noted, no significant relationship was found between JOL and test performance across all three tests. Table 7 provides the correlations for JOL and test performance.

Midterm

The writing and reading comprehension test scores were less variable compared to the other two sections. For the writing section, the lowest score was 0 ($n = 2$), and the highest score was 97, with an average score of 66.36; for the reading section, the lowest score was 46, the highest score was 100, and the average score was 79.45. The writing and reading sections of the test were highly and significantly correlated, $r = .80$, $p < .01$.

For the midterm writing section, correlations were as follows: preparation and writing test score showed positive correlation, but was not significant, $r = .25$; knowledge and writing test score was negatively correlated, $r = -.21$; difficulty was positively correlated at .48, and confidence was negatively correlated with test scores at $r = -.60$. For the skills section, knowledge, difficulty, and confidence JOL were negligibly correlated with skills test scores, $r = .01$, $r = -.11$, $r = -.14$, respectively. Reported preparation for the test was negatively correlated with test scores, $r = -.38$.

Writing Test 2

The average score for the in-class writing was 62.91, with the lowest score a 0 ($n = 3$), followed by 77, and the highest score a 97 ($n = 1$). Students' reported preparation for the test was

negatively correlated with their writing scores, $r = -.39$; knowledge JOL and test scores were not correlated, $r = .02$; difficulty was correlated negatively, $r = -.35$, and confidence $r = -.26$.

Writing Test 3

The average score for this writing test, 62.18, was similar to the previous test. The lowest score was 0 ($n = 3$), followed by 70, and the highest score was 97 ($n = 1$). Test scores for this writing test were moderately correlated with knowledge (.35), and difficulty JOL (.43); confidence and preparation were minimally correlated with test scores, $r = .29$ and $r = -.19$, respectively.

Table 7

Correlations between JOL and Test Scores for Linda's Students

JOL	Mid-Term			
	Writing ($n = 7$)	Reading ($n = 10$)	Test 2 Writing ($n = 8$)	Test 3 Writing ($n = 8$)
Writing				
Preparation	.25	-.11	-.39	-.19
Knowledge	-.21	-.16	.02	.35
Difficulty	.48	-.54	-.35	.43
Confidence	-.60	.33	-.26	.29
Skills/Reading				
Preparation	-.47	-.38	-	-
Knowledge	-.28	.01	-	-
Difficulty	-.26	-.11	-	-
Confidence	-.35	-.14	-	-

Note. * Significance at .05 level; ** significant at .01 level

Summary of RQ1 Results

To summarize, knowledge ratings showed significant positive correlations with test scores for the three unit tests; difficulty, preparation, and confidence ratings, however, did not show any consistent patterns in their relationship to students' test scores. Even though knowledge ratings were highly correlated with preparation and confidence ratings, only student

ratings on how well they knew the material related to their performance scores. JOL and performance correlations for Linda's class were consistently negative; in other words, JOL ratings were lower for students with high test scores and vice versa. However, even with the presence of high correlations, none of the associations were statistically significant.

What is the Difference between Student Metacognitive beliefs at the Beginning and End of the Semester?

This question was addressed by conducting paired sample *t*-tests of student responses in the pre- and post-student surveys. The metacognitive beliefs scale was comprised of eight items, each item a statement related to student's beliefs about his/her writing and reading skills, $\alpha = .90$. The response options were organized along a five-point agreement Likert-type scale, where 1 = strongly disagree and 5 = strongly agree; the scale included items like "I have a good understanding of English grammar" and "I can summarize what I read in English." The survey included two other scales that were also analyzed using paired-sample *t*-tests : Influences on Student Writing and Application of Classwork. The influences on English writing scale included five-items, also with a five-point Likert-type agreement scale ($\alpha = .57$). The items related to different factors influencing a student's writing, such as "Reading my classmates' writing helps me improve my own writing" and "I am good at writing in my native language." The application of classwork scale was a two-item ($\alpha = .70$), five-point agreement scale that included the following items: "When writing in English, I use what I learned in class"; "I apply ideas from course readings to in-class activities like discussion or exercises."

Prior to analysis, the data were examined to ensure that the absence of significant outliers in the difference scores and the assumption of approximate normality in the distribution of the difference scores were met. Difference scores were computed, and outliers were examined in the

frequency distribution. One student participant selected 1 (strongly disagree) for all items in the post-survey scales. As a result, the student's scores stood out as outliers and affected the distribution of the scores. Skewness and kurtosis statistics showed no significant skew in any of the difference score distributions, but there was a significant kurtotic effect on the metacognitive beliefs scale, influences on writing, and the perceived difficulty of reading item (6.45, 6.30, and 5.80, respectively; standard error of the kurtosis statistic was 1.12, 1.12, and 1.09, respectively); since the statistic was more than three times the standard error, the difference scores distributions were deemed excessively kurtotic. The student's data were deleted, prompting the researcher to re-examine assumptions of normality. Two of the three scales indicated normal kurtosis statistics. For the perceived difficulty of reading item, there was another outlier where a student's perception of reading difficulty changed from 3 (difficult) in the pre-survey to 1 (not at all difficult) in the post-survey. Since this anomaly was not due to a data error and represented a valid data point, the case was retained in the paired sample *t*-test.

Table 8 provides descriptive statistics on the pre- and post-survey responses and results from paired sample *t*-tests. Fifteen students completed both surveys, and one student completed the pre-survey and a portion of the post-survey (See Table 5 for details on sample size). Therefore, the sample for perceived difficulty of the four English language domains was $n = 16$; for all other scales, $n = 15$.

There were no significant differences in students' perceived metacognitive beliefs, $t(13) = -1.98, p = .07$ (Pre: $M = 3.36, SD = .860$; Post: $M = 3.66, SD = .601$), influences on writing, $t(13) = .11, p = .92$ (Pre: $M = 3.70, SD = .357$; Post: $M = 3.69, SD = .507$), or application of coursework, $t(13) = .49, p = .64$ (Pre: $M = 3.79, SD = .611$; Post: $M = 3.71, SD = .508$).

For perceived difficulty of English language skills ($n=15$), responses ranged from 1–4 (1=not at all difficult, 2=somewhat difficult, 3=difficult, 4=very difficult). Although there was a slight increase in the mean difficulty, there was no significant difference in perceived difficulty of reading skills, $t(14) = -.37, p = .7$ (Pre: $M=2.07, SD=.594$; Post: $M=2.13, SD=.743$). There was no difference in perceived difficulty of writing skills, $t(14) = -.00, p = 1.00$ (Pre: $M = 2.60, SD=.632$, Post: $M = 2.60, SD = .507$). Students reported a significant decrease in difficulty of listening skills, $t(14) = 2.26, p < .05$ (Pre: $M = 2.40, SD = .986$ Post: $M = 2.13, SD = .990$). Similarly, students also reported a significant decrease in perceived difficulty of English speaking skills, $t(14) = 2.26, p < .05$ (Pre: $M = 2.40, SD = 1.056$ Post: $M = 2.13, SD = 1.125$).

Table 8

Descriptive Statistics and Paired Sample t-test Results of Pre- and Post- Student Surveys

	Metacognitive Beliefs about writing		Influences on Writing		Application of Coursework		Perceived Difficulty							
	Pre	Post	Pre	Post	Pre	Post	Reading		Writing		Listening		Speaking	
<i>N</i>	14	14	14	14	14	14	15	15	15	15	15	15	15	15
<i>M</i>	3.36	3.67	3.70	3.69	3.79	3.71	2.07	2.13	2.60	2.60	2.40	2.13	2.40	2.13
<i>S</i>	.860	.601	.357	.507	.611	.508	.594	.743	.632	.507	.986	.990	1.06	1.13
<i>D</i>														
<i>t</i>	-1.98		.11		.49		-.37		.00		2.26		2.26	
<i>p</i>	.07		.92		.64		.72		1.00		.04		.04	

Qualitative Results

The following sections describe categories and themes that emerged from participant interviews and observations. The sections are organized in the following manner: The first section provides a brief overview of the coding and analysis of the qualitative data, which includes six teacher interviews and eight classroom observations. The second section describes themes that developed from each category including contextual factors, teacher and student characteristics, and formative assessment practices. The third section attends to the synthesis of results as a means to address the following research questions: What are the ongoing formative assessment practices of teachers in ESL classrooms? (RQ3) How do teachers in ESL classrooms use classroom assessment to inform instruction? (RQ4) How do ESL teachers use data on metacognitive judgments of learning in their formative assessment practices? (RQ5) To understand how the results speak to these research questions, however, one must first examine the procedures for analyzing the data informing said results.

Coding and Analysis Procedures

The primary approach to coding and analysis was to use the theoretical framework of formative assessment informing the study design and research questions. Themes and categories were extrapolated from the data to identify the formative assessment practice used, given the context and its participant characteristics. However, important descriptions of class activities and teacher beliefs on instruction and assessment were coded using a grounded approach. Categories represent “code families,” or overarching codes that frame individual codes or themes. Themes are significant topics that emerged from the coding and analysis; they were determined based on the frequency of a code or a group of related codes and/or theoretical significance. Figure 4 provides a summary of the coding and analysis process.

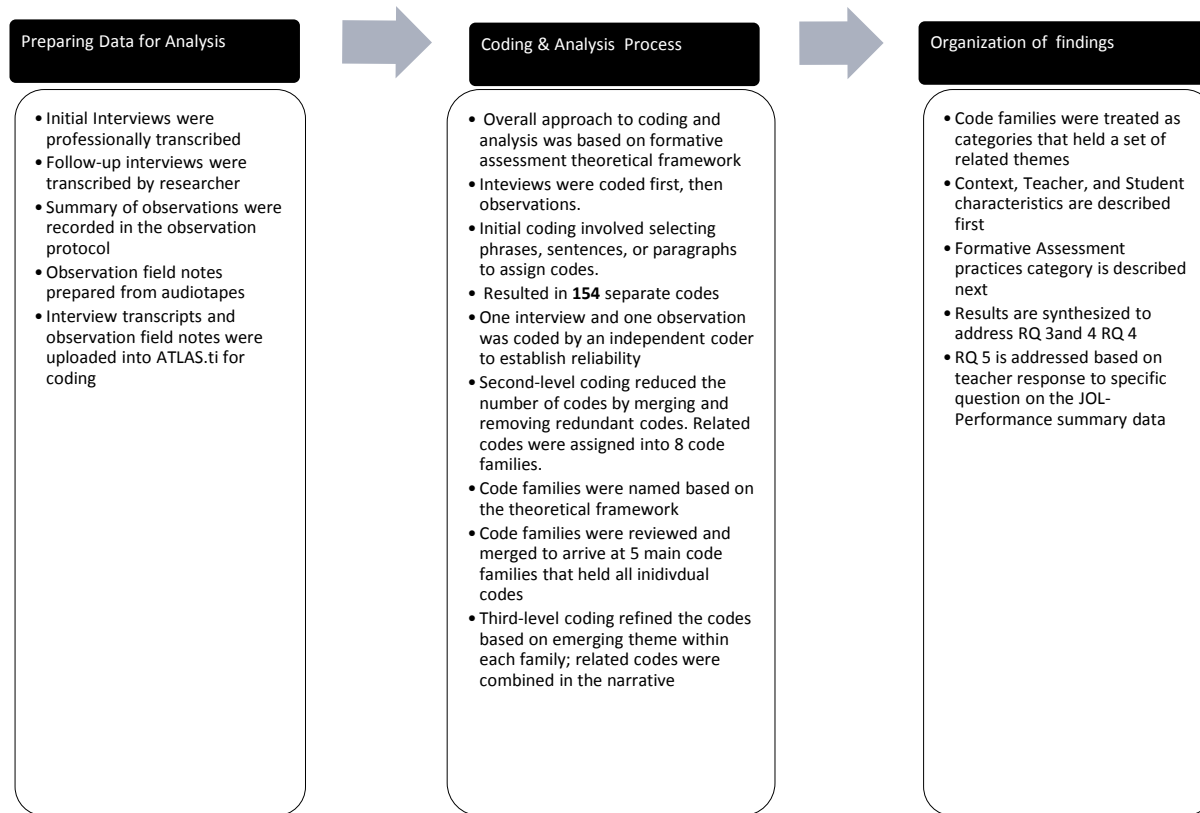


Figure 4. Summary of the coding and analysis process. Description of the coding and analysis procedures employed for the interview and observation data.

Coding process. The three initial interviews were transcribed by a professional transcriptionist, with the three follow-up interviews being transcribed by the researcher. The researcher also prepared detailed field notes for the eight observation sessions and recorded a summary of activities in the observation protocol. The observation field notes and the interview transcripts were color-coded by participant to easily attribute quotes. The text files were uploaded to ATLAS.ti for coding and analysis. Each transcript and set of observation field notes was first read in its entirety, and then re-read line by line to initiate coding. The texts were revisited a third time to ensure consistency and assign new or previously overlooked codes. The rationale for such an iterative approach to coding was to provide consistent and thorough

descriptions for understanding contextual and individual (teacher and student) factors in formative assessment practice.

The initial coding of data was informed by sample coding guidelines put forth by Bogdan and Biklen (2007), and Lewins and Silver (2007). Bogdan and Biklen described categorizing codes according to setting, situation, perspectives, activities, processes, etc. Similarly, Lewins and Silver presented inductive and deductive approaches to coding based on a theoretical or grounded approach. In addition, in their discussion, Lewins and Silver provided coding and analysis strategies for identifying emergent themes using ATLAS.ti. Both sources (Bogdan & Bilken, 2007; Lewins & Silver, 2007) emphasized that there is no one way to develop codes, and that a multifaceted approach to coding was acceptable, if not preferable. The use of different approaches simultaneously is attributable to fact that the coding process evolves as the researcher makes sense of the information. A total number of 154 codes were created during the initial coding step (See Appendix G for the initial and the final list of codes).

The first set of interviews was coded first, followed by the follow-up interview transcripts and observation field notes. Codes were single words or phrases that represented the essence of a line, sentence, or interaction. Descriptions of activities and participant perspectives on instruction and assessment were coded using an inductive approach. These codes were not informed by a preassigned conception or theory; rather, they were assigned as they occurred, and were revisited upon a second and third read-through. Using a deductive approach, a list of codes related to formative assessment practices was developed from the interview, observation protocol, and relevant literature. Formative assessment practice-related codes (sharing expectations, using assessments, making instructional adjustments, and feedback codes) were developed from the framework of formative assessment theory. The development and analysis of feedback codes, on

the other hand, were informed directly by the observation summary sheet. For example, the observation protocol was used to determine feedback-related codes such as the specificity of feedback (individual, group, class), feedback type (positive, corrective, metacognitive), or nature of feedback (checking student understanding, elaborating on student understanding, etc.). In order to enable the triangulation and validation of interview data, feedback-related codes were also constructed so as to reflect language similar to that used to code the interviews. All observation codes started with the prefix “Obs” to distinguish them from interview codes. After all documents were coded, the codes were then analyzed to identify strategies for reducing the overall number of codes and for further detecting themes.

Inter-rater agreement. A primary aim of the coding procedure was to ensure the thoroughness of the coding process, as well as contribute to the confirmability, credibility, and dependability of the inferences made from the analysis (Krefting, 1991; Soslau, 2012). Meeting this goal entailed soliciting the help of an independent coder to establish inter-rater agreement. In order to establish inter-rater agreement, the coder, who was familiar with higher education classroom settings, and well-versed in qualitative research, coded one interview and one observation. Prior to coding, the researcher and coder reviewed the list of codes containing definitions and example quotes from interviews and observations, at which time, any questions and concerns raised by the independent coder were addressed by the researcher. After the researcher and coder had reviewed 50% of the first interview, they proceeded to discuss the coding and discrepancies. Following this discussion, the second half of the interview was coded and a similar discussion was initiated. Inter-rater agreement was calculated by the percentage of codes (in relation to the total number of codes) used by both coders. The first half of the interview resulted in 50% agreement, and the second half, 38%. The inter-rater agreement for

the observation field notes was 60%. Discrepancies in coding can be explained by the independent coder's use of umbrella codes versus the researcher's use of more specific codes. There were instances where the coders assigned different but related codes to a text, which was indicative of redundancy in codes that were later merged for analysis. To elaborate, instructions for codes like "educational qualification" and "teaching experience" were not clearly provided; hence, the coder assigned two or more instances where only one occurrence per transcript was expected. The lower inter-rater agreement for the second half of the interview may be attributed to the occurrence of several new codes that were not addressed in the first discussion.

Mismatches in coding could also be attributed to differences in experience and knowledge related to the study of formative assessment in ESL contexts. For example, the coder's use of the broad code of "assessment practices" versus the researcher's use of informal and formal formative assessment codes like "informal FA: checking student understanding" illustrates this point. Additionally, there were 58 codes that the independent coder had to study prior to coding, thus suggesting her unfamiliarity with such codes; the coder's unfamiliarity with certain codes may have contributed to her use of more general and easily identifiable codes, such as "assessment practices" or "classroom interaction."

Although inter-rater agreement was low to moderate, a discussion between the researcher and coder indicated that a majority of the discrepancies identified could be resolved during the next phase of coding and analysis when redundant codes would be merged, or when codes would be subsumed under one general code or category. Where there were clear discrepancies that indicated the code was inappropriately assigned, such codes were removed from further analysis. Both raters thus reached a negotiated agreement that informed the second level of coding and analysis (Soslau, 2012). In qualitative research, the concept of reliability is not necessarily

supported by the consistency of two individual raters' coding, but by the degree to which their findings from an identical context are consistent (Bogden & Biklen, 2007). While the two coders may not have reached the desirable inter-rater agreement, they consistently identified similar codes for texts and phrases.

A second-level coding procedure was performed where redundant codes were merged, and irrelevant or marginally relevant codes (< 5 occurrences) were removed. Specific codes were also merged to form a more general code; for example, assessment and grading practices, as well as attitudes toward assessment and grading, were merged into the single code of assessment practices. Following the merging and removal of certain codes, the resulting codes were classified into eight major coding families, including: marginal codes, activities, setting, context, strategies, teacher characteristics, student characteristics, and formative assessment. Several codes were assigned to two or more code families, and often to multiple related families; for example, the "feedback strategies" code was assigned to both the strategies as well as formative assessment code families. Similarly, many codes classified under setting were also coded to context. In order to address these redundancies, setting and context code families were combined, and codes included in the marginally relevant code family were removed after determining that they were not relevant to the research questions. The product of combining, condensing, and eliminating codes was five code families, viz., contextual factors (codes related to the description of the setting, influence of the program/department, teacher collaboration, flexibility related to assessments, etc.), class activities (descriptions of activities), teacher factors (teacher beliefs, education, experience), formative assessment (feedback strategies, nature and specificity of feedback, sharing learning expectations, making instructional adjustments), and student characteristics (codes related to student behaviors like tardiness, absences, motivation,

class participation, cultural influences, academic expectations, etc.). The class activities category was merged with the formative assessment category during analysis. The reorganization of codes and code families facilitated the emergence of themes, as code families themselves were considered categories for classifying main themes. Table 9 provides an overview of the categories and themes derived from code families.

Categories and Themes

The following section describes formative assessment as it is situated among contextual, teacher, and student characteristics. Formative assessment practices are described in the context of the three research questions, with each teacher's individual practice woven into the narrative as a means for comparing, contrasting, and highlighting significant points.

Table 9

Description of Categories and Themes for the Coding and Analysis of Interviews and Observations

Category, Theme	Codes*	Description/ Example Quote (Source)
Contextual Characteristics	6	Themes related to the setting/context of the study
Classroom Setting		My classes are in a lab, I've always tried to schedule them in a computer lab so students can be writing on the computer right in the classroom (Interview 1- Thomas)
Assessment/Grading Policy		<p>"The emphasis is heavily weighted toward writing as you can see, 50 of the grade comes from the portfolio" (Interview 1- Thomas)</p> <p>"(curricula objectives are)...set by the office, the department, the administration..." (Interview 1- Daniel)</p> <p>"I was given a lot more freedom, but I had to put together a curriculum along with materials that we are matching the same curriculum..."</p>
Teacher Characteristics	15	Themes related to specific teacher-related factors like education, teaching experiences, and beliefs related to instruction and assessment
Education & Experience		Qualification, teaching experience in ESL, and training in assessment

Attitude toward assessment and grading		For me, to give a 100 on a writing assessment would mean it has to be perfect. That means no grammar mistakes, all the content is there, your organization is perfect (Interview 2: Linda)
Frustration		I send out comments on the draft and the next draft will come back and some of things have not changed. It might be something simple like <i>capitalize the title appropriately</i> and it is not done... and it's frustrating. (Interview 2: Thomas)
Student Characteristics	15	Includes themes related to student factors like culture and/or classroom behaviors
Culture & Academic Experiences		a lot of (student success) has to do with what they expect and it also has to do with what type of culture, if you don't come from a reading/writing culture, it's hard to sit down and read. (Interview 2: Linda)
Successful & Struggling students		I have example after example of students who come regularly, take the suggestions that I offer, or that come from the book, from me, from my experience, they implement those into their work and show ... huge amounts of improvement from day 1 to ... week 16 (Interview 2 – Daniel)
Class Participation		For general questions, the teacher asks the class to respond. For specific questions related to the exercise, like going over the five questions, he calls on students (by name). When the question is open to the class, only 2 -3 students are active in responding. (Observation 1 – Thomas – Section 2)
Formative Assessment Practices	51	Themes related to the process of formative assessment
Sharing Expectations		We start with the objectives, we learn what the objectives are and then we read how to attain the objectives and then we go back and we get the examples... (Interview 1- Daniel)
Informal Assessment as Evidence		Often times the questions they ask will show that they didn't grasp what the assignment was, what was expected of them (Interview 1 – Thomas)
Formal Assessment as Evidence		... formal assessment ... helps me to determine the next thing to work on (Interview 2- Linda)
Feedback Strategies		My comments are not, you know, I don't fix it for them. They are pretty vague, I want them to figure it out. But I would expect them to ask me that. (Interview 2 – Thomas)
Student Use of Feedback		I'm not sure a lot of times the students are reading my feedback because I will see the same things being repeated, the same errors... (Interview 1 Thomas)
Instructional		I plan to give feedback but then when there's a gap in

Adjustments	understanding, that's where you just change something completely. Which is why I have that two weeks, and I always go behind and then I catch up. (Interview 2 – Linda)
-------------	--

Note. * texts or phrases may be assigned to two or more codes.

Contextual Characteristics

The first interview included the following question to gather information about program influences on instruction: “To what extent are your assessment and grading practices influenced by expectations and curriculum set by the English Language Program?” The same question was asked during the follow-up interview, and the teacher’s response was summarized. After asking this question in the follow-up interview, the researcher continued by asking the teacher if anything happened during the semester that would change his or her response. While responding to other questions in the protocol, teachers also frequently alluded to program policy, or grading and assessment policy, and described their overall attitudes toward teaching, grading, and assessment. Observation field notes included descriptions of the physical classroom to gather information on its role (if any) in formative assessment. Together, these data sources guided the analysis of the teaching and learning context.

Physical characteristics of the classroom. Thomas taught two class sections in the same classroom. This classroom was a large lab, and each student desk had a computer. In the front of the classroom was a big projector screen connected to the instructor’s computer. The lab had three rows of seating for approximately 25 students, and each desk had side panels to provide privacy. The monitor was placed on top of the CPU on the desk. The observation field notes noted that the structure of the monitor and CPU “hindered the view of the instructor and other classmates.” The setting was ideal for workshop-style individual work and lecture, and was less suited for group activity and class discussion. The students typed notes and took tests on the

computer during class. Additionally, in all observation sessions, the teacher provided hand-outs for practice exercises.

Daniel's classroom was artificially lit, and featured a whiteboard and projector screen connected to the instructor's computer. The room was small, with individual desk chairs that could seat a maximum of 12 students. The chairs were not arranged in a particular order; rather, they were pushed to the edges of the classroom. The teacher used the whiteboard to write sentences during grammar lessons and to note class announcements. He used the projector screen to display class material for students from Blackboard or email. The students completed in-class writing tasks and unit tests on paper.

Linda was in a classroom similar in size, seating, and ambience to Daniel's classroom; however, the wall facing the student seats was "marker-friendly," which enabled the teacher and students to use the whole wall as a whiteboard. In addition, the classroom had a whiteboard on an easel and a projector screen that operated from the instructor's desktop computer in the front of the classroom. The teacher used the wall and the easel whiteboard extensively during all observed class sessions. Every review activity featured in the class was visual in the sense that the teacher demonstrated errors and used color markers to make corrections or to direct students' attention to distinctions. Students completed their writing tasks and tests on paper, but they also wrote their responses to practice exercises on the wall prior to reviewing them as a class. In the follow-up interview, Linda added that sometimes the class had enough laptops that enabled students to type their responses and share it with the teacher; she would then review the activity on the projector screen instead of the wall.

Assessment and grading policy. The assessments for WC 3 were comprised of writing and reading components, with greater emphasis on writing. One assessment, the writing

portfolio, which was comprised of a collection of essays and commented-upon drafts that represented student work throughout the semester, accounted for 50 of the course grade. Reading skills were assessed by objective-type unit tests, as well as midterm and final exams. There was a midterm and final exam, which accounted for 20 of the final grade, and four unit tests amounting to 20 of the final grade; participation and homework were allocated 10 of the course grade.

All class outcomes and objectives were determined by a curriculum set by the program. In their interviews, teachers indicated that the program administration set the curriculum and its associated objectives; hence they strongly influenced teachers' assessment and grading practices. However, teacher committees, including committees for section teachers, lead teachers, and skill coordinators, drafted assessments and established the grading scale. The lead teacher for a course created the tests for all sections, except for the Community-Based Learning Written Communication 3 course (CBL WC 3). In constructing assessments for their classes, individual teachers had the option to adopt an 80-20 approach, where they could add or change 20 of test content. The CBL WC 3 teacher, who was not required to administer a common exam, was granted the flexibility to determine both course materials and assessments for her course, as long as they still met course-wide learning objectives. However, all other components of the regular WC 3 grading scale were applicable, including equivalents of unit tests (in-class writing and reading tests) and writing portfolio requirements.

The main objective of the WC 3 course is to provide students with English writing and reading skills that would enable students to take on undergraduate (or graduate, in some cases) academic coursework. In the first interview with Thomas, he described how the goals of the ELP bear upon his teaching practice, remarking that "the goal and objective of the English Language Program is to assure that students' English language skills are proficient enough to carry full

time academic classes. So that being the goal, I feel that in my assessing and passing students I'm saying that they're ready for full-time academic classes" (Interview #1). Students who failed WC 3 had to repeat the class before they could proceed to take regular university classes. Passing the class also marked the successful completion of the ELP curriculum.

The program set certain rules like not failing a student based on attendance, and that performance would be the main criterion to determine a student's grade or whether he or she passed or failed. Student attendance was an issue that all participants identified as affecting learning and instruction. Daniel, for instance, felt he did not have to do anything about it because students who did not come to class were likely to fail the class anyway on account of poor performance; he noted, "The office hands down all these rules and regulations...but the point is, you know, a lot of students just don't come to class and don't do the work...then they don't pass. It's that simple"(Daniel, Interview 1).

Even though the ELP set the standards for attendance and the criteria for passing or failing a student, teachers nonetheless enjoyed flexibility in determining the assessments that they chose for the class; although teachers said the program had a major influence on their assessment practices, they emphasized that it was teacher committees that determined the course content, grading scale, and assessments. As Thomas stated in his first interview, things like performance criteria, grading scales, and assessments are

pretty much set by the teachers. Again, all sections of teachers have to agree on that, so it's a discussion among teachers. Now there are some rules and things we're told by the (office), for example, we cannot fail a student based on their attendance, for example, I mean, there are some rules that we must follow.

In this way, teachers did not feel that they did not have a voice or the authority to implement instructional and assessment activities since they were developed by them in collaboration with other ELP teachers. This sense of teacher autonomy was especially pronounced in Linda's interview, where she describes it as an advantage that she was able to put together a course packet for the CBL course that matched her teaching style. In response to a question on how much time she spent planning for each class, she said, "it might be a little bit less (than other classes) because the materials I got (that) I was able to choose (for the class),... are all books I have used before, so I had already seen the materials before" (Interview # 2).

Teacher Characteristics

As teachers discussed issues like those mentioned above in their interviews, the following themes emerged from questions related to teachers' assessment and grading practices, as well as personal reflections on their teaching style and classroom experiences. The first section of the interview asked questions pertaining to participants' educational backgrounds, training, and experience in assessment.

Education and experience. All three teacher participants held master's degrees with a specialization in teaching English as a foreign or second language. They have taught ESL reading, writing, speaking, and listening courses. Thomas has 21 years of teaching experience; Daniel has 17 years; and Linda has 20 years. Participants also took on non-teaching roles within the program; for example, Thomas served as the skills coordinator for the three levels of Written Communication, and Daniel acted as the lead teacher for Written Communication.

All three teachers took one or two classes on assessment during their graduate study, yet Thomas and Daniel rated their training in assessment to be minimal; Linda, on the other hand, rated her training in assessment to be moderate to extensive. Thomas, the lead teacher for WC 3

classes, had developed the assessments for the current semester. Speaking to his experience with assessment, Daniel indicated extensive experience in developing and using rubrics for writing assessments. Additionally, Linda had work experience in test development and standardization for ESL contexts. Despite their limited formal training in assessment, all three participants perceived their skills in assessment development to be strong or very strong. Along these lines, Thomas noted,

As teachers you think...oh, you just make a test based on what we've covered, and that is it. But designing a test is a real skill, it's a real skill. And I feel I've gotten better, but I still feel that I make mistakes. Designing a test to measure what you want to measure is challenging. (Interview #1)

Linda demonstrated a similar appreciation of assessment development when she said that she “love (s) developing (tests) and hate(s) grading papers” (Interview #2).

Teaching, grading, and assessment style. While all three teachers expressed their attitudes toward teaching, assessment, and grading in their interviews, their views toward each subject were not necessarily uniform. Thomas, for instance, tried to be flexible in determining final grades, preferring to consider a student's participation in class and discipline in addition to performance on tests, as opposed to fixed final scores at the end of the semester. On this subject, he reflected, “Evaluating a student is more than testing. That's a component and an important one. But you have to look at other factors too” (Interview #1). At the same time, he put in place clearly established rules and expectations and followed a similar routine in both sections.

Thomas described himself as rigid in his teaching methods; he felt students perceived him to be a “tough teacher, a difficult, a demanding” (Interview #1) teacher, but he hoped they viewed him as fair. Observation field notes indicated that he had a formal, yet friendly, tone of voice, and

that he was consistent and predictable across his interactions with students. Even in his feedback to students on their writing tests, he patterned his interactions with different students more or less the same.

Contrary to Thomas, Daniel characterized his teaching style as social and relaxed, which, he noted, worked for some students, while others took advantage of this laid-back approach. In observation field notes, he is portrayed as frequently sharing his personal experiences with his students and as adopting an informal tone while conversing with them; furthermore, the students appeared comfortable interacting with him, albeit in conversations unrelated to class content. The following extract from field notes for Daniel's first class observation recounts a discussion about whether or not the class should meet the next day, owing to a religious holiday observed by the students. The teacher expressed his feelings of frustration about tardiness with students while maintaining an informal, social tone:

- Student 7:** Tomorrow, there is class tomorrow?
Teacher: (with some irritation) That's a good question. You show up 40 minutes late... we can talk about class tomorrow based on your performance today.
(Student 7 giggles)
Teacher: How would you rate your performance? 10 being the best, and 1 being the worst- what would you rate your performance?
Student 3: Eight and a half.
Teacher: Eight and a half? You were late too.
Student: Just 10 minutes
Teacher: Just 10 minutes. That's late too. (The teacher narrates a personal story that influenced him not to be late to class) So, I am not giving you a break by helping you be late. So, what do we do next? I say we just quit and go home (with frustration). Is that what you want to do?
(No response)
Teacher: What about tomorrow? Are you guys going to come tomorrow?
Student 7: Yes.
Teacher: You? You are going to come?
(Student laughs)

Consistent with Daniel's more relaxed teaching style, he reflected on the subjective nature of grading writing, asserting the need for rubrics that allow for creative differences. He preferred to frame his assessment "around building on what you know, not comparing with others" (Interview #1). Moreover, he saw the writing portfolio, which comprised 50% of the course grade, as an opportunity to balance grades by considering whether or not students met the requirements of the rubric in conjunction with their improvement over the course of the semester. In continuing to describe his assessment practices, he placed particular emphasis on peer assessment, where he provided minimal feedback that only included covering procedural aspects of completing a writing assignment (showing an example, and providing practice):

Until their peer review group understands what (students) are trying to say... I don't really need to see it because I am not going to correct it for (them). When (they) give it to me, I am going to give them a grade. (Interview #1)

He added that this strategy makes students work together and take peer assessment seriously.

Unlike Daniel, who initially provided minimal feedback to students, Linda said she "believe(d) in immediate feedback ... in them seeing their own mistakes" (Interview #1) because she felt that students only looked at the grade they received at the end of an assignment, not the teacher's comments. Even though she used assessment tools like worksheets, she did not grade them and instead asked students to share their work in a group as a learning exercise. She also believed in grading student work based on what she would expect from them at the end of the semester: "For me, to give a 100 on a writing assessment would mean it has to be perfect. That means no grammar mistakes, all the content is there, your organization is perfect" (Interview #2). This meant students received lower grades on their writing at the beginning of the semester. While students often found such high standards difficult to handle so early in the semester, Linda

believed in “a lot of positive reinforcement” (Interview #2), and in constantly communicating her grading style to help them understand course objectives. Linda was reflective about her instructional style, remarking that she aimed to frequently connect her methods to lessons learned through previous teaching experiences. She believed taking the time to cover basic material in the beginning of the semester was worthwhile because grammar and reading skills are acquired sequentially. She distinguished herself from other teachers by following a slow pace at first to ensure students had learned, and even though she may have fallen behind in covering material, she indicated that she was able to always catch up in the end.

Frustration. Thomas and Daniel expressed frustration at students’ behaviors, including missing classes, and not doing classwork or homework. In characterizing their frustration, Thomas and Daniel cited students’ tardiness, poor attendance, and lack of attention in class. Thomas noted that in his second section there were four or five students who were not engaged, and “it sucked the air out of the class for me and the rest of the students” (Interview #2). He found that students were preoccupied with their phones, and this led him to ask students to place their cellphones on a side table for the duration of the class, a rule he enforced throughout the semester. During the observations, he instructed students to turn off monitors prior to a lecture, review activity, or announcement. Both Thomas and Daniel, however, were observed expressing their disappointment to students in class. For both teachers, homework review was hindered by the fact that very few students completed the homework. During observation of his second section, Thomas expressed his frustration to students when a class activity did not go as planned because of non-participation from students:

Boy, you gotta read, you gotta read... if you are not doing the homework you are not getting the experience in reading and finding answers... that is the exact kind of questions

you will get on the tests. So, doing the homework is practice... and the other questions, you have to be able to read, that is the source to finding where the answers are.

During the follow-up interview, Thomas described a day when his frustration led him to end the class abruptly.

One of the first things I would do in homework review (is ask) how many of you did the homework? and in one section, sometimes maybe one or two students would have done the homework. One day, I was just so upset, I said-Ok just go, and I just ended the class.

Daniel experienced something similar in the first observation session, when he was following up on an assignment and asked each student if they had completed an outline for an essay and brought it to class. Most students replied they had not completed the assignment or that they did not bring it to class. The teacher expressed further disappointment when a student asked for more time to complete the outline, even though it had been assigned several weeks ago. During the first observation of this class, the teacher said to himself aloud that he did not know what he was going to do that day if students had not completed the homework. He described his frustration in the follow-up interview in this way: "I just gave up on homework. If I did give homework it was very general."

Student Characteristics

The category of student characteristics is comprised of the most frequently marked codes, with a total of 199 quotations assigned to 15 codes. Several themes related to student characteristics emerged from the interviews and observations. One such theme refers to teachers' acknowledgments of the influences of culture and educational or academic experiences on student behavior and performance. More than culture, also included in the category of student characteristics, are teachers' descriptions of students' academic or educational experiences.

While student educational experience is not a direct result of their culture, it does seem to be closely related to a student's background. Students' academic or educational experiences and their individual characteristics, like their desire to learn or motivation, were singled out by teachers as the most important factors of student success. Another theme represented in this category stems from teachers' descriptions of successful and non-successful students; these descriptions were offered in response to interview questions related to characteristics that the teachers associated with students who received As in the course, and those who received Fs. Related to these attributes, a final theme that emerged from teacher responses and researcher observations was "class characteristics," or the collective composition of students in a class, which affected a teacher's experience. In addition, teacher responses as to whether and to what extent students used formative feedback, as well as observations of student response to feedback in class, added to the researcher's understanding of the role of student characteristics to the formative assessment process. The following paragraphs present the above themes, in detail, from the perspective of the broader category of student characteristics.

Role of culture and academic experiences. As English language learners in a foreign country, students' cultural backgrounds and experiences influenced their behaviors in the classroom. In fact, all three teachers noted that while a typical characteristic of struggling students was that they were not active participants during class, there were also exceptions; for example, some students were not necessarily struggling in class, but were "more reticent... shy" because of a cultural background where students are "trained to listen to what is given to them, absorb it, and learn that way" (Thomas, Interview 1). Linda described some students as not wanting to "lose face, so they won't ask questions in front of others" (Interview 2), which

prompted her to consider allocating a special time outside class for each individual student to address questions about class content.

Linda felt that students who were on government scholarships may not have taken their education as seriously as students who bore the financial burden of their education themselves. Further, she also attributed age as a factor causing students pursue to “having fun” rather than their studies, as they were experiencing a culture in the United States less restrictive for young adults than their own. In combination with lax program policies on attendance, these students pursuing fun do not face repercussions or see incentives to attend class, and hence struggle to complete the course. Government-sponsored students were allowed three attempts to pass each course. As a result, students who fail repeat the class two or three times before exiting the course, as Daniel described:

A lot of my students are repeats, I tend to get the ones who have given a couple of shots at this class and I’m kind of the end of the line... could say I am the remedial teacher, pick up the ones who either don’t get it or don’t try...by the time they get to me, they have already been through this, they already know the routine, they already know the deal. There is a certain amount of arrogance too, you know, ‘I know this,’ ‘I took this.’

Cultural or national background also influenced how students communicated with each other during class. Since a majority of students were from the same country, students typically conversed in their native language with each other, even as they engaged in a class activity. The researcher noted several instances where Thomas and Daniel encouraged students to talk to each other in English or explicitly asked them to refrain from talking in their native language during class. Linda addressed this issue by purposefully assigning students to groups where no two

students shared the same native language. Because her class was more diverse, this strategy was a more feasible and successful solution.

All three teachers referenced the role of students' educational or academic experiences in their class performance. Performance was related to their cultural background, as well as individual students' willingness to thrive in an academic environment. The following comments speak to an alignment between a student's previous academic experience and the academic expectations of an ELP within the American higher education system. Thomas indicated that whether or not students used feedback depended not only on their motivation to learn, but also on their "cultural educational practices or things they've learned and experienced in how they've performed in classes in their own country and their own culture, whether or not they were encouraged or whether or not they felt comfortable (seeking and using feedback)" (Interview #1). Similarly, Linda believed students' educational and academic experiences to be the foremost factor of student success: "it doesn't have to do with ability as much as it has to do with their expectations and their previous experiences and their willingness to then become an academic student" (Interview #1). Additionally, at another point in the interview she said, "A lot of it has to do with what they expect, and it also has to do with what type of culture, if you don't come from a reading/writing culture, it's hard to sit down and read." When referencing her group of students, she defined struggling students as "in over their heads, sometimes they don't understand the expectations of academics, whereas I have a strong group...eight out of twelve who definitely understand the dynamics of education and what's expected" (Interview #1). Daniel likewise attributed student success or failure primarily to individual student characteristics. He said,

Two students the same age, same majors, different countries, one took suggestions, came to every class; the other one was just arrogant and did not even consider what I was saying made any sense, did not show any improvement and the one who came to class, same age, same majors, same level, showed just extraordinary improvement in every aspect... it could be personality too... it's not country, religion or race oriented.

(Interview #1)

The next theme pertains to attributes of successful and failing students as they are documented in interview responses and class observations.

Attributes of successful and failing students. In addition to demonstrating competence in reading and writing, Thomas portrayed students who received an A in the course as disciplined, punctual, regular, and prepared for class; they did the homework and contributed to class discussions. Daniel described students who received an A grade as taking learning seriously.

They're long-term oriented. I think they see that the reason for being here is to further their education and they're willing to accept challenges, and let's face it, writing is not easy and reading some of this stuff is not easy. I think also, the ones who are more serious minded, I mean, it's okay to have fun, but it's also ... to have fun learning the difficult information. So the ones who realize and come to class; ... attendance plays a big issue as well... I have example after example of students who come regularly, take the suggestions that I offer, or that come from the book, from me, from my experience, they implement those into their work and show vast huge amounts of improvement from day 1 to ... week 16. (Interview #2)

For Linda, successful learners were the ones who did homework, were motivated, and

understood the academic obligations that make them good learners.

Students who struggled or failed the class were presented as the opposite of the successful learners and were distinguished by their not attending class, coming to class late, and not doing their homework. All participants noted that students who put forth effort showed improvement; none of the teachers attributed failure to lack of ability. Daniel claimed that the student who does not see improvement is the “one that’s been skipping and failing all the tests, the unmotivated, apathetic ... spoiled brat, who’s here for the party” (Interview #2). He added that

if the student is really giving it a try and attending class, and working with others, and really trying to understand the concepts, that’s the student that will come to me or come to the professor, the teacher, and ask, what’s going on here, I just can’t quite get this.

Linda felt that students are not inclined to take learning seriously when they do not see repercussions for their lack of success; this absence of consequence reflects lenient program policies, such as those related to attendance and tardiness. Linda also added that students who struggled with the course material showed a lack of academic skills and failed to help themselves, but that they realized these shortcomings. With respect to students’ lack of surprise or concern about their unsatisfactory performance in the course, Thomas provided an example of a failing student who came to him toward the end of the semester to ask if he was likely to pass. When Thomas reviewed the student’s performance with him and said it was not possible that the student would pass, he anticipated needing to defend his remarks. However, the student accepted Thomas’s remarks and simply went on to ask if he should attend or skip the final exam.

Class characteristics. Another theme originating from the observations dealt with class participation, and how teachers’ instruction was influenced by the collective makeup of the class.

As mentioned previously, Linda's class was distinct from other WC 3 sections in that students met outside of class for a substantial amount of time to satisfy the course's community service learning component. Linda felt that not being confined to a traditional classroom setting may have helped students bond and work together in class, as the class even went to lunch on two or three occasions. In the first interview, Linda admitted that the class dynamics were not typical for the context; she believed that students entered the class with traits that made them successful. She expected most students to pass the course with an A or B, and attributed their success to motivation and knowledge of academic expectations. Observation field notes described this class in a consistent manner, noting that this class was distinctly more engaged than the other three sections. Students were active and responded to the teacher in a way that demonstrated interest and engagement. Although two to three students were the most active participants, other students contributed to discussion and review activity more than students in the other three sections. Illustrating this point, the following excerpt is a reflection from the second observation of this particular class:

Just as last time, I find the class to be lively; students seem interested and engaged, and they participate in classroom conversations. There is laughter and informal off topic conversations which I noted in another classroom (Daniel but not in Thomas). I feel like there is a sense of community perhaps because students get to work with each other outside class during service learning. They seem more comfortable working with students who are not their native language speakers, and there is more movement and activity.

Unlike Linda's class with its heterogeneous demographic composition, Thomas had two sections that were similar in demographics, as the majority of students hailed from one country. Research observations indicated similar dynamics between the two classes, although one class (section 1)

had 17 students, whereas the other class (section 2) was smaller, with 12 students. Moreover, the teacher's methods and class activities were similar for both sections. He followed a routine of starting the class with a timed practice exercise, followed by reviewing the exercise as a class, then a lecture/new material, followed by exercises pertaining to lecture content; if there was time remaining, students sometimes worked on writing assignment drafts. The teacher was also consistent in his method of reviewing exercises. For each item, he called out a specific student to provide the correct answer. When the student responded correctly, he moved on to the next item; when the student was unable to answer, he opened up the question to the rest of the class. He followed every response, correct or incorrect, with a repetition and elaboration of the student response. Student responses were minimal and restricted to responding to teacher prompts or questions. In terms of responding to teacher-initiated discussion, only two or three students were active participants; the rest of the class looked at the monitor on their desk or the textbook. Observation field notes identified a couple of instances where, recognizing the class' lack of participation, Thomas expressed his frustration.

Even though from outward appearances the classes seemed to behave similarly, during the follow-up interview, Thomas characterized the second class as being difficult, owing to disruptive classroom behaviors, and this behavior seemed to affect his motivation.

The two classes were quite different. The first class was much better overall... but the second class had a group of students may be 4 -5 that just weren't engaged and it just sucked the air out the class, both for me and the rest of the students... there were days when I didn't look forward to that second class. And I realized that wasn't good. I mean, it just pulled me down.

In her observations of this section (section 2), the researcher noted a couple of instances that

exemplify the teacher's point. One observation note reads, "Two students who seem to be athletes walked in at around 11:45 am (45 minutes late); they take the last row by me and both put their head down for the rest of the class." Among other disruptive classroom behaviors, Thomas added he had to take steps to prevent students from using cellphones in class by asking them to put them on a side table for the duration of the period. He continued that he did not know what caused the differences between sections, wondering if it was merely a matter of chance that his two classes were so different from each other.

Daniel's class differed from both Linda and Thomas's class in that he had only male students, and all but one student belonged to the same nationality. During her observations, the researcher noted that this class had more absences than the others, with only 50% of students present on both observation days. The ones who did attend participated in informal conversations with the teacher, but were less active and engaged during class discussions. Only one student consistently and actively responded to questions and clarified doubts with the teacher during lecture and review. Daniel, in the follow-up interview, stated that this was not a strong group and students were "routinely unprepared for class," which often led to the students working on the homework in class. He attributed the lack of participation to the class time (9 am), and possibly to some students attempting to pass the course a second or third time. Although he expressed his disappointment, he believed about half of the students who came to class and did the work showed improvement at the end of the semester.

The following category includes emergent themes related to formative assessment.

Formative Assessment

Data from teacher interviews and classroom observations were analyzed based on the different attributes of formative assessment, viz., how teachers share learning expectations, how

they elicit evidence of student understanding, and how they provide feedback and/or make instructional adjustments to support student learning. The following themes represent their own sub-sections, and are organized according to the process elements of formative assessment: sharing learning expectations or goals; eliciting evidence of students' current learning from formal and informal assessments (including peer assessment techniques), making sense of the information; and taking action to close the gap between current understanding and the learning goal.

Sharing learning expectations. All teacher participants were asked to describe if and how they communicated learning goals to students, followed by whether they thought students typically understood the learning goals. *Sharing expectations of learning goals* is interpreted as instances where teachers refer to the goals of the class, a concept, a class session, or a test. This reference could be made in relation to a specific activity, or it could be made at the more global level, in relation to class goals. All teachers recounted sharing learning expectations for the course at the beginning of the semester and periodically during the semester. Lending support to teachers' claims, observation notes detailed several instances where teachers relayed the purpose of an assignment or activity to students; that said, teachers did not, however, interpret "sharing learning goals" in the exact same way.

In the first interview, Thomas claimed that he communicated learning objectives by being clear about the grading scale and how students would be graded on both the reading and writing components. He said while he talked to students individually about what they needed to do to pass the class, the syllabus contained global learning objectives. He felt it was important to share learning expectations, and thought that something more formal and explicit than what was provided in the syllabus should be given to students. He reiterated this point in the follow-up

interview, commenting that not much had changed in the way of class objectives during the semester: Reading skills were still primarily assessed using objective-type tests; writing assessments were still graded using a rubric. In the follow-up interview, he said he went over the rubric with students and urged them to learn the expectations prior to a couple of writing tests.

In her class observations, the researcher recorded five instances of the teacher sharing learning expectations with students. In the first observation for section 1, Thomas told students that the readings and tests were going to get difficult because they are expected to be able to cope with intro-level undergraduate courses, which “have a lot of reading”; He continued by warning that “it’s going to be difficult reading and you need to understand if you want to get through these courses.” In addition to sharing expectations for course material and assessments, Thomas set students’ expectations with respect to the day’s activities. At the beginning of each class, he would verbally apprise students of his plans for class as a way to communicate day-to-day learning objectives. On one occasion, Thomas even provided a handout containing the day’s activities to students. While observing section 2 (Observation 2) toward the end of the semester, the researcher witnessed Thomas spend several minutes going over the writing portfolio requirements, and explaining the purpose and contents of each component.

Like the others, Daniel said that he goes over the objectives in the syllabus at the beginning of the semester. He also said that before covering new material, the class read the learning objectives associated with each unit/chapter in the textbook, followed by the objectives for the end-of-unit assignment. Remarking on this approach, Daniel said, “We work, really, a lot of times from the inside out rather than in a chronological way” (Interview #1). He viewed sharing learning expectations as “extremely valuable” in that “the student knows what’s expected of him” (Interview #1). However, he admitted that in the beginning, students do not always

understand learning expectations, but that they gain confidence over time. The teacher also spent the majority of a class session going over the elements of the writing rubric prior to a unit test (Observation 1). He read the description for the maximum points possible for each element, and reminded students that it was the same rubric that was used for the previous writing assignment.

As for Linda, she said that sometimes the objectives listed in the syllabus “could be a little over (students’) heads” (Interview #2). So, she used this as a learning opportunity, displaying the syllabus on the projection screen and asking students if they understood the content listed in the objectives (e.g., thesis). She admitted that students do not always understand the learning objectives, and that it is challenging for some students even after going over objectives in class or having peers explain them. Her strategy to help students internalize objectives was to grade student work according to said objectives. In the beginning it was difficult for students to cope with lower grades on account of that, but she constantly used verbal reinforcements and pep talk to remind students that her grading practices were guided by her final outcome expectations.

Informal assessments as evidence of student learning. Teachers described the different ways they gathered information on whether or not students understood course content. In the classroom, all teachers used questions, for example, asking students if they have questions, and treated the nature of questions and students’ level of participation as indicators of student learning. Teachers were observed using questioning as a way to check student understanding in at least 40 instances. They posed questions to the class about content covered or to gauge student understanding of content they were about to cover. They also asked students if they had any questions at the end of a review or instructional unit.

Thomas also asked students to repeat what they heard as a way to monitor student

understanding. He paid particular attention to how well students understood the instructions or requirements of an assignment for the reason that “often times the questions they ask will show that they didn't grasp what the assignment was, what was expected of them” (Interview #1). The aspect of formative assessment noted here may be challenging to English language learners who are using the language they are learning to follow the instructions or expectations of an assignment. Thomas implicitly addressed such a challenge by asking a student to repeat the requirements of an exercise or assignment, and then elaborating on the student's explanation. It is likely that a student's performance on a class activity or an assignment/test reflects whether or not he or she understood the instructions. In the event that a student performs poorly on an assignment or assessment, and the reason behind the performance is left unexamined by the teacher, it may be interpreted as a deficit in content comprehension. On attributing poor performance to not understanding the instructions as opposed to the content, Thomas conceded that he found “that sometimes after the fact that oh, they didn't understand, or they saw the assignment as something completely different” (Interview #1).

Daniel used different methods for gathering evidence of student learning, including giving mini-quizzes and monitoring small group activity. Furthermore, small class sizes (and low attendance) enabled him to talk to individual students during class. In consulting with students one on one, he found that some students' learning was hindered because they were using books with previously marked answers in them. The fact that students offered right answers in class led him to mistakenly believe that their responses were indicative of learning. It was only when his students performed poorly on a test covering material learned in class that he realized that they had, in fact, not learned. For him, informal assessments in the form of small group activities and daily interactions helped inform instruction, but he relied more on formal assessments to gather

evidence of student learning. In both observation sessions, Daniel's class engaged in lecture and small group activity for most of the class. Both sessions focused more on skills like creating an outline for an essay, reviewing the writing component of the unit test, and identifying test-taking strategies for the reading component. Given that the classes worked primarily on the writing component of assessment, the researcher did not get to observe Daniel's techniques for checking student learning with respect to reading skills. Yet, in one observation session, Daniel did check for students' understanding of the rubric for the writing component by asking them questions and elaborating on student responses to reiterate the content of the rubric.

As for Linda, she said that every once in a while she collected evidence of learning via writing assignments, and by gauging student questions and reactions during class activities. Like Thomas, she too asked students to paraphrase or rephrase what they heard, and sought to include students who were quiet or had many absences during review activities. She also considered students who participated and regarded their misconceptions as informative sources for evidence of learning. She listened in on conversations during small group activity to gauge students' understanding, and approached each group to see if they had questions. During observations, she asked students to raise their hands to "vote" on answers to a question she had posed. She used this technique when she did not receive a response to her initial question, but wanted to quickly gain understanding of what students knew and increase overall participation.

Let's vote. If you think 'because' goes (in this part of the sentence), if it is one of the subordinators, raise your hand. Be proud, be bold. Everybody has to vote. If you weren't here yesterday, guess. If you think 'because' is one of these (connectors) raise your hand. Be proud, be bold. This is how we learn. Let's do it again. 'Because' goes over here (connectors) raise your hand; 'because' goes over here (subordinate) raise your hand- it

goes here (insubordinate) - remember Mina was here and she said, “yeah, Americans, we make that mistake all the time”? One of the most common fragments - because I love you ... what? (Observation #1)

Class activities as formative assessment opportunities. Information on class activities was obtained from the two observation sessions for each class and a specific interview question; the interview question read, "Could you describe a typical day in a WC3 classroom?" During the follow-up interview, participant responses to this question were summarized, and teachers were asked if there were any changes to their initial responses. Responses in the follow-up interviews were consistent with observed class activities. Class activities were modified based on teachers' judgments early in the semester regarding students' strengths and weaknesses. The following paragraphs describe these class activities that formed the foundation for formative assessment practice in this context.

Thomas described a typical day as one that began with five minutes of free writing as a warm-up exercise, followed by review of homework, presentation of information (lecture), and then practice exercises on the topic covered. The remainder of the class was dedicated to in-class writing, during which time students continued working on writing assignments for their portfolio. In the follow-up interview, Thomas made changes to this routine because students showed clear weaknesses in the reading and skills sections of the unit tests early in the semester. For both sections, he replaced the five minutes of free writing with a TOEFL practice reading exercise. To develop students' reading skills, every day, he started the class by handing out a reading exercise; he set a timer on the projector for 10 minutes as students worked individually. Once they were done, students consulted with a partner to check answers, after which the teacher

reviewed the exercise as a class. This change to the daily routine remained in effect for the rest of the semester.

Daniel described a similar routine comprised of lecture, followed by individual or small group practice exercises. He followed the book closely in determining class content and practice exercises. In the follow-up interview, he said that he changed the focus of the class from writing to reading skills based on talking to other teachers about students' low performance in the reading section of the first two unit tests: "I poured more emphasis into the skills, and the practicing of the skills" (Interview #2), relied on exercises in the textbook, and had students work in small groups.

Linda began class by discussing the homework due next class and proceeded to give practice exercises for reading skills that could be performed in small groups, which were then shared on the "wall" for discussion as a class. She incorporated unplanned activities based on what students were struggling with, and anticipated considerable changes in terms of how much she covered each day. She planned her classes for the week, and even though the routine varied day-to-day, she would get through most of her weekly lesson plan.

All three participants cited homework review as a typical day's activity. Linda felt that performance on homework was an indicator of a student's success in the course: "if you do the homework, you tend to be successful" (Interview #2). Similarly, Daniel said, "75% of learning in college takes place at home, it's homework" (Interview #2). As a result of the importance placed on homework, both Thomas and Daniel set aside class time to review homework. Homework was not graded, but contributed to a class participation grade that totaled 10% of the final grade. As noted in the observation field notes, class sessions for all sections included either a review of

homework or explicit instructions related to homework due the next class. Teachers were observed giving reading exercises, readings, and writing exercises (e.g., outlines) as homework.

Although teacher participants assigned practice exercises as an individual activity via homework, they also had students work on exercises in groups. Such exercises were typically reading or grammar exercises comprised of objective, multiple choice, or short answer items. Practice exercises were mentioned in participant interviews as part of a typical day's activities, but observations indicated they were actually a major feature in all three classes, taking up most of the class time. Students engaged in various exercises related to the day's lesson or to previously taught content at the individual, small group, and class levels. In Daniel's class, students practiced writing skills in small groups, whereas in the other two classes, teachers and students engaged in reading exercises one on one.

There were certainly differences, however, in the extent to which each teacher incorporated individual, small group, and class activities in the classroom. Thomas used individual practice activities, where students corroborated answers with a partner if time permitted. This was consistently followed by reviewing the exercises as a class. Daniel, on the other hand, stated in his first interview that he used peer or small group activities extensively and also relied on peer assessment for feedback on student work. Consistent with what he stated in his interview, observations note that Daniel mostly assigned practice exercises and writing activities to small groups; these groups worked together and submitted one product for a class participation grade. Daniel provided a rationale for his preference for group work by saying, "I do small group work; I think it's called cooperative learning where everyone in the group receives the same grade...so that way I throw the burden of responsibility of teaching the class on the class" (Interview #1).

Linda used small group activity extensively in class sessions. She emphasized that she preferred informal assessments, and while working in small groups, she is able to see what students do and do not understand. She purposefully assigned students to work with partners who did not share their native language, insisting that “no two speakers of the same native language can be in the same group” (Observation #1). She also grouped students based on skill level and personality. In order to manage time, she assigned specific items from an exercise to each group. Once they finished working on the exercise, group members shared the responsibility of writing their responses on the “wall” to prepare for class discussion.

Formal assessments as evidence of learning. In addition to class activities, all three teachers regarded performance on the unit tests and essay drafts as important indicators of learning. Thomas, for example, used unit test results to gauge student understanding at a class level; he stated,

For example ... collocations, which are words that go together frequently, I remember, students did so poorly on the test, and I couldn't figure out why they did (be)cause they were things we just went over in class and right out of the book, and it made me realize that they didn't understand it well. (Interview #2)

Much like Thomas, Daniel utilized unit test results to come to the conclusion that students in his class generally had a good grip on vocabulary and grammar. Additionally, as noted before, it was due to students' poor performance on unit tests that Daniel realized that they were responding correctly in class only because they had used textbooks with previously marked answers in them. He also attributed the ease with which he was able to identify what skills students were missing in their writing to his decades of teaching experience. While daily interactions helped guide the

class' direction and signal areas where students needed to improve, formal assessments helped him separate the serious students from ones who were just there for fun:

[Through] writing assessments, or in class assignments, you can see who's growing, who's developing, and who is still having underdeveloped content, unspecific support, evidence that doesn't support. So it's pretty obvious... who is doing (the work) and who isn't. (Interview #2)

Linda said she used more informal ways to assess students, including ungraded exercises and small group activities, because her class was so motivated that "they didn't even need (grades)" (Interview #2). She clarified that the worksheets she used in class were indeed assessment tools, but she found them to be more valuable when they were treated as interactive and reviewed in groups or as a class, instead of graded. She added that formal assessments were useful in writing, though. She used a color-coding technique to make corrections, using green for grammatical errors, and purple for content-related errors. This system indicated the group's strengths and weaknesses, as she noted when she said, "My group here has a lot of purple... they need to work more on organization and content" (Interview #2).

General Feedback Strategies

During the first interview, teachers were asked about their feedback practices in their daily classroom interactions and on unit tests. The follow-up interview included a similar question, which asked participants if they would like to change or add to their previous response. The observation field notes recorded teacher-student interactions to analyze teachers' feedback practices. The following section describes their feedback practices based on the insight gained from the interviews and observations.

Thomas did not typically give feedback on student performance in the skills/reading

portions of the unit tests; for the writing portion, however, he did provide individual feedback to students during one-on-one conferences. In her observations, the researcher noted one class session that was dedicated to providing verbal feedback to students on their writing by holding individual conferences as the rest of the class engaged in a practice exercise. The researcher observed five of these interactions, and each interaction adhered to a pattern that Thomas identified in the follow-up interview as a strategy that he used for drafted essays in the portfolio as well.

I work on grammar in one paragraph, usually on one paragraph maybe a little more... so I'll make comments there. In the end notes, I will ask them to look at those, correct anything that (they) can correct, and then read the rest of the paper and look for similar types of errors they are making. I think helpful comments are the ones that help students identify the kind of mistakes they are making so they can focus on them.

As students worked on subsequent drafts of each essay in the portfolio, his comments on their writing became fewer and shorter. He believed giving too many comments could be counterproductive: "My comments are not, you know, I don't fix it for them. They are pretty vague; I want them to figure it out. But I would expect them to ask me that" (Interview #2). Daniel framed his feedback on student writing in a similar way.

Rather than taking an essay, and writing all over it, I filtered them back with minor comments, then last couple of weeks of class, I've sat down with each student individually, and gone over and I have had them rewrite it. I don't wanna see a draft 1 or draft 2, bring me a draft 3- you go to writing center and help each other. So a lot of times when they sat down with me, they've got the focus, organization, content... what we are dealing with for the most part is grammar. (Interview #2)

He preferred not to mark extensively on student writing after he realized from past experience it “look(ed) terrible” (Interview #2) and may have discouraged students. Therefore, he refrained from providing corrective feedback on students’ writing, and instead relied on global feedback related to content or organization. He expected students to work with peers or seek help at the writing center to make revisions. He preferred to let students take in the feedback, and discussed students’ writing only if they wanted to follow-up on the information.

Linda’s feedback on student writing comprised of questions and positive remarks. On assignments and tests, she made more formal edits to content and organization. She also described looking at student essays to address gaps in understanding. For example, in one of the observations, Linda made a worksheet out of anonymized sentences from student essays, and then asked students to improve the sentences. She created the worksheet after reading students’ essay drafts and realizing that they had trouble framing topic sentences. Using students’ own work in this way functioned to make activities relatable to students and provide them with direct peer feedback. She also observed a pattern of errors related to fragments and run-on sentences among student essay drafts, and thus decided to revisit identifying and correcting run-on sentences, even though she had already covered this topic.

Feedback-related interactions. A primary goal in observing and recording teachers’ feedback techniques was to gather information on the techniques that best complemented their formative assessment practices. A sample of teacher-student interactions was transcribed verbatim and coded to analyze feedback practices. The codes included the main codes of feedback type and feedback specificity. Feedback type included corrective feedback, where a teacher acknowledged whether or not a student response was correct with a yes/no reply; positive feedback, where a teacher replied to a student response with a positive comment, as in “very

good”; and metacognitive feedback, where a teacher’s response to a student’s question or answer, instead of simply providing corrective feedback, prompted students to think more. Specificity of feedback was organized into sub-codes that included details about the nature of feedback, viz., whether it was given to the class, group, or individual student. Codes also accounted for whether feedback was linked to learning goals, elaborated on a student response, if it was a question framed to gauge student understanding or a suggestion to make improvements to a response.

Teachers’ feedback to students varied in nature, specificity, and frequency. Referring to literature on feedback (Hattie & Timperley, 2007; McManus, 2008; Panova & Lyster, 2002), teacher-student interactions were coded according to feedback type—positive, corrective, metacognitive, or linked to learning goals—and feedback specificity—individual, group, or class, checking student understanding, elaborating on student understanding, and suggestions for improvement. As far as feedback types, corrective feedback was most frequently used by Thomas and Linda. Questions directed at students by teachers were most commonly used to check student understanding; teachers frequently asked students if they knew the meaning of a word or phrase, and reviewed previously covered material before proceeding to a new topic or exercise. For example, while reviewing a practice reading exercise that students had just concluded, Thomas asked a series of questions related to the items to check whether students knew the strategy for reaching the correct answers, like, “Where in the paragraph do you think you will find the theory?” or “What would you look for in the first paragraph?” He repeated each student response as it was offered and reiterated the concept related to the response. He called on specific students, and if they did not respond or gave an incorrect response, he asked the class to respond. His immediate feedback was corrective or positive, followed by an elaboration of the

response before proceeding to the next item.

Figure 5 illustrates how review interactions took place in all three classes. Teachers asked questions to elicit student responses in order to check understanding. When student responses were appropriate, teachers either moved on to the next item, or, as in many instances, they repeated the answer and reiterated the concept behind the response. Student participation was an important determinant in these interactions; typically, only one or two students actively participated in all classes. Others passively followed the activity unless specifically called upon to respond. This was a consistent pattern observed in all four sections. Student responses also triggered reactions from teachers. At times, no response from students caused Thomas and Linda to comment on the difficulty of the exercise, or express discontent or frustration.

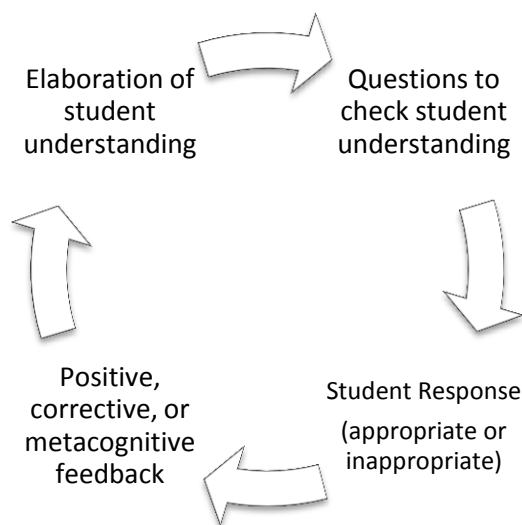


Figure 5. Illustration of observed feedback related interactions. An illustration of feedback-related interactions observed in all three classes during a review activity.

Table 10 provides the frequencies of the feedback interactions illustrated in Figure 5.

Table 11 provides extracts from the observation field notes for all three teachers that illustrate these interactions.

Table 10

Feedback Practices of Teacher Participants Coded from Observed Teacher-Student Interactions

Feedback	Thomas*	Daniel	Linda	Total
Type				
Corrective	8	2	5	15
Positive	4	0	4	8
Metacognitive	4	1	1	6
Specificity				
Individual	12	3	2	17
Linked to learning goals	7	1	1	9
Check student understanding	15	9	16	40
Elaborate student understanding	10	5	13	28
Suggestions for improvement	5	1	2	8

Note. * Code counts for Thomas represent occurrences over four observations in two sections.

Table 11

Examples of Feedback -Related Interactions for the Three Teacher Participants

Feedback -Related Interactions for the Three Teacher Participants		
Thomas ^a	Daniel ^b	Linda ^c
<p>What is a suffix? Student 1: end of the word; Teacher repeats their answer and elaborates that it is short Student (female): changes the part of speech. Teacher repeats.</p> <p>Teacher asks for an example. Students give several like <i>al, tion</i>. ...</p> <p>Teacher: paraphrases the reading and asks what do you call something that comes in the beginning of the word Student (female): Prefix Teacher reads out the different examples of suffixes given in the book and then starts the exercise on suffixes.</p>	<p>Teacher moves on with the rubric-grammar and mechanics – use of pronouns. He says he does not like the use of pronouns and prefers the use of repetition of use of keywords to keep the reader aware and engaged about the subject being talked about. He repeats this a couple of times for emphasis. He asks students what is a pronoun, and one of the students replies “he/she”, and the teacher responds affirmatively, and elaborates with a definition – it replaces a noun- the use of Starbucks, and the associated pronoun for it. He reiterates that pronoun use, especially use of a wrong pronoun makes the writing unclear, and says it is better to use Starbucks repeatedly to keep it clear.</p>	<p>T: complex. And an example of simple? S3: Life is funny (other students laugh) T: (laughs) Life is funny. You guys are great... you remembered it. And I know I said not to use exclamation points much, but this one needs one. Okay, a compound sentence? S2: Life is fun, but it is fun to die. T: What was Gabriel’s example? S: I love you, but I want to be alone. T: And then complex sentence?</p>

a. Observation 1, section 1; b. Observation 1; c. Observation 1;

Instructional adjustments. Instructional adjustments refer to modifications to instruction as observed by the researcher or described by the participants that were the result of formative assessment. Thomas claimed that if students were unable to demonstrate their learning in class, more needed to be done to help them understand. He said he revisits test sections or items where most students performed poorly. If only some students performed poorly, he felt compelled to continue without re-teaching in the interest of time and to “get through the material in the book” (Interview #1). In the second interview, he noted that students performed poorly on the reading/skills portion of the tests, which led him to speak extensively about the instructional changes he planned to make for the following semester. For the current semester, though, he provided more in-class reading exercises. For example, in the first interview, he described a typical class day as one that started with five minutes of free writing. However, after learning of students’ poor performance on the reading/skills sections of the tests, Thomas introduced ten-minute reading exercises to help students strengthen their skills. Reflecting on this change in the follow-up interview, he said students were divided on whether they liked the modifications. For the next semester, he planned to implement both free writing and reading exercises, or alternate between them. Even though these exercises “take a good chunk of time,” he claimed to be open to such a change since it would be “to the best advantage of the students” (Interview #2). Also in the follow-up interview, Thomas described major changes he planned to make to future WC 3 assessments as a result of students’ poor performance on reading/skills assessments during the semester. He felt that as they were, the assessments were heavily focused on writing, so he was contemplating changes that would make the assessments more equally focused on both reading and writing.

Daniel made similar changes to his instruction. His focus was typically on writing, but

during the semester, it became apparent to him from unit test scores that students were struggling with the reading and skills portion of the class. He said, “Whereas in the past, I’ve tried to focus more on the writing and work reading in, this semester I focused on reading and worked writing in, and with the more dedicated students, it actually seemed to work really well” (Interview #2). Daniel talked to other teachers of the same class who voiced similar concerns, which led him to increase the focus of instruction on reading/skills and incorporate writing. He added that the change had worked as intended, and based on the writing portfolios he received, he saw improvement.

Linda was more reflective about instructional adjustments she made during the semester. As noted earlier, she depended heavily on informal assessments to determine instruction. She said she was able to change instruction based on the areas where students were struggling. She provided several examples of changes she made to the content she covered or revisited, which included cutting short or extending instruction based on students’ knowledge of the content. She described her process for making instructional adjustments in the following way: “I got that from the worksheets and seeing what they were doing, and what they knew. And it’s like okay, you guys got this; probably what they need the most is vocabulary development and paraphrasing” (Interview #2). Her plan for providing feedback and revisiting topics was flexible. She used formal assessments mainly to help her identify areas where students needed additional instruction, which, in turn, allowed her to use the next instructional unit to cover such areas, instead of revisiting the test results with the class. She anticipated and accounted for these adjustments in her plan for the semester by setting aside the last two weeks of class to make up for any delay in covering the required content. In the first interview, she emphasized that managing time while making the necessary changes to the course schedule can be challenging.

As an example, she said, students requested her to cover more grammar even though she believed they did not need additional grammar instruction. She wanted to give them more exercises, but that would put a strain on the class's pace: "little exercises, they become long explanations (that) take out maybe 20 minutes".

Student use of feedback. The interview protocols also included a question on how teachers knew if students were using the feedback provided to them. While teachers did not provide definitive answers to this question, they drew connections between use of feedback and student characteristics like level of seriousness, or other attributes that make them successful. When students did not use the feedback given to them, the teachers became frustrated, as Thomas noted:

A lot of times... I 'm not sure...the students are reading my feedback because I will see the same things being repeated, the same errors, the same mistakes being made. And sometimes when I give feedback on writing and the students will turn in the next draft, it's almost the same draft they turned in the last time. So it's disappointing and it's discouraging. Some students do show progress and make changes, but it's a slog.

(Interview 2)

According to Daniel, successful students were characterized by their tendency to take suggestions from the teachers, and these were the students who saw improvement over the course of the semester. Linda felt her set of students took feedback, and if they did not internalize or make changes in response to feedback, it was on account of not understanding the feedback, rather than neglecting to consider it altogether. She strongly believed that students needed to see their own mistakes, and she enabled this through peer assessment and other techniques, such as color- coding errors or comments on student work. She connected student

response to feedback to students' academic expectations in that students who cared about learning adopted behaviors that fostered it, which included taking and acting on feedback. She also reiterated the importance of class dynamics where the positive students set the tone for the rest of the class, thus persuading students who typically do not take the class seriously to participate and contribute.

Observations provided insight into the nature of student responses in teacher-student feedback related interactions. In a formative assessment situation, appropriate responses indicated students provided correct responses, inappropriate responses indicated incorrect or an unrelated response, and no response indicated that students did not respond to a teacher question. There was a fair amount of all three types of responses observed by the researcher; however, appropriate responses were the majority (25), followed by no response (18), and few instances of inappropriate responses (10). These codes are not indicative of class participation because it is possible that one or two students were responding to teacher questions more frequently than others. Another code, student-initiated formative assessment was used to record instances where students initiated a formative assessment or feedback interaction in class. In total, there were 12 such instances, of which 5, 3, and 4 to students of Thomas, Daniel, and Linda, respectively. All the quotes under this code related to students asking questions or doubts to teachers in the middle of a review activity (individual, group, or class) which triggered a feedback interaction as illustrated in figure 5.

Synthesis

Teachers varied in the formative assessment techniques they implemented; Linda demonstrated the most variety in her techniques; class activities provided ample opportunity in all three classes for formative assessment. Teachers' techniques were influenced by their own

style and beliefs, program expectations, as well student attributes. All three teachers demonstrated making instructional adjustments based on assessment information, but they differed in the nature and extent to which formal classroom assessments informed instruction. The two research questions related to the ongoing formative assessment practices and the use of formal classroom assessments that were the focus of the investigation are addressed below to synthesize the themes and categories. Following that, analysis and findings on the third research question related to how teachers used the metacognitive learning judgments data they received are presented.

What are the ongoing formative assessment practices of teachers in ESL classrooms?

Teachers' formative assessment practices took place within the background of their teaching and assessment experience, their attitudes toward assessment and grading, and program expectations and policy. All participants were qualified, highly experienced teachers, who were confident about their assessment development and instructional adjustment capabilities. Thomas and Daniel believed to have received minimal training in assessment, whereas Linda reported that she had extensive training in assessment.

Their practice of formative assessment consisted of a variety of attributes; together, the teachers demonstrated that they shared learning expectations with students at the global level as well as at a task level. Their approach was different though –Thomas conceded that a more explicit way to communicate learning expectations is needed even though he was observed task based sharing learning goals; Daniel valued the need for sharing learning expectations and recognized students initially found expectations difficult to understand, but his methods seemed procedural or like a formality (for example, reading the syllabus, reading the goals for each lesson in class); in light of the difficulty students face in understanding learning goals, Linda

used the syllabus as a learning opportunity and used constant communication with her students to keep them from feeling discouraged when she graded their initial drafts based on final expectations.

All three teachers used questioning as a method to elicit evidence of student understanding of content. A peculiar aspect of this technique was questioning to check whether students have understood the instructions to assignments and tests. Perhaps, doing so ensured that a student's performance on an assignment was not on account of failure to follow the assignment or test requirements. In another form of questioning, teachers, especially Thomas and Linda, relied on the nature of the question students asked in class as evidence of student understanding. All three teachers engaged in a feedback cycle that provided corrective and positive feedback and elaboration of student responses to questions; while Linda and Thomas were observed using this technique more frequently, there were fewer instances recorded for Daniel whose class sessions during observations were more focused on writing activities.

The teachers were also different in their instructional style, which influenced their formative assessment practices. Thomas, for instance, adopted a rigid teaching style even as he engaged in feedback techniques maintaining a formal tone with students. His techniques were primarily teacher directed, and this was most evident as he reviewed practice exercises. Contrary to Thomas' teaching style, Daniel's was social and flexible, but he depended more on peer assessment and global-level feedback on student writing. As garnered through observations, he spent a substantial amount of class time going over the writing rubric to prepare students for the writing component of the unit test; his verbal feedback interactions were consistent with the principles of formative assessment, but were less varied and frequent. Linda was flexible in her teaching style; she used a variety of techniques to elicit student understanding, and to provide

formal feedback. She made continuous instructional adjustments based on what she learned about where students were struggling. She demonstrated comparable flexibility by changing her assessment plans based on student need, and used informal assessments (non-graded assessment tools) extensively in class.

The teachers reported having flexibility when it came to designing and using assessments. Although assessments and grading scales were standard across classes, they were developed by teacher committees, and Thomas and Daniel played an active role by taking on non-teaching administrative roles related to curriculum and assessment development.

Program expectations did influence formative assessment practices to an extent: All participants thought attendance was a major concern, especially because poor attendance was an attribute of students who struggle. Yet, they could not fail a student based on attendance, nor was there any policy in place to encourage or require students to show up to class. In the follow-up interviews, teachers expressed frustration at students not coming to class or coming unprepared for class because it frequently disrupted their instructional routine even leading teachers to dismiss class. Thomas's classes had high attendance compared to Daniel's class, which had about 50 attendance on both observation days. Daniel said he liked to treat students as adults in college, and he felt if a student did not attend class, but was able to pass assessments, it was acceptable; though, he added that such an occurrence was rare. Similarly, Linda said such cases where students were able to succeed despite poor attendance were truly exceptions, and that typically students who did not attend failed because they missed learning opportunities in class.

The teachers displayed a level of helplessness in their responses related to the above issue, but they dealt with it in different ways. Thomas said he did not want his disappointment to affect his instruction and demeanor for the students who do show up and want to learn; hence, he

chose to ignore some student behaviors but also imposed restrictions where he felt they were needed (as in the example of turning computer monitors off and leaving cellphone at a side table for the class duration). Linda maintained an attitude of being happy that an irregular student showed up at all and tried to involve him or her in class activities.

Figures 6, 7, and 8 summarize the three teachers' formative assessment practices in relation to significant teacher, student, and contextual characteristics that emerged from the interviews and observations. Among these factors, student characteristics, including attendance, tardiness, academic experiences/expectations, and class participation were cited by all participants as determinants of student success or failure. Several examples from interviews and observations demonstrate how these factors affected teachers' instruction. Students' final grades are included in the figures to provide additional information on the class characteristics. While final course grades are the result of multiple factors, including performance on assessments, participation, and teachers' grading styles, examining the attributes of successful and non-successful students in relation to grades, as well as the class composition, can help make sense of the climate within which the teachers engaged in formative assessment. For example, Thomas and Daniel expressed frustration at student behaviors including tardiness, attendance, coming unprepared to class, disruptive classroom behavior, and non-participation in class activities; many of these are attributes that all teachers listed as characteristic of unsuccessful students. Thomas said that these behavioral issues, especially in section 2, demotivated him. Almost half of the students in Thomas' section 2 (46 or 6/13) failed the class, and only one student received an A, compared to section 1 with 35 (6/17) of students earning an A and 18 (3/17) failing the class. Daniel's class was similar in that 27 (3/11) of students failed the class, and the majority passed with a C average; only one student received an A.

In Linda's class, all students passed with an A or a B, and three students failed the class. These three students were described during the follow-up interview as being out of the loop throughout the semester: "I have three or four that won't pass that is because they didn't do any of the work and they stopped coming" (Interview #2). She considered this class to be atypical in their high level of motivation and engagement. For her, students who wanted to work hard and were aware of academic expectations did what is needed to succeed in a class, whereas students who were not successful often had excuses for not taking class seriously. She also clarified that students came in with positive attributes, which helped create a successful learning experience: "The students who were successful that is ... 9 out of 12 ...their awareness of what they needed to learn, how they needed to learn it, what they needed to do were up there. They had it all" (Interview #2). Even though the majority of her students exhibited positive characteristics, Linda nonetheless believed that it was very challenging to instill motivation in students, a difficulty she was experiencing with another class she was teaching.

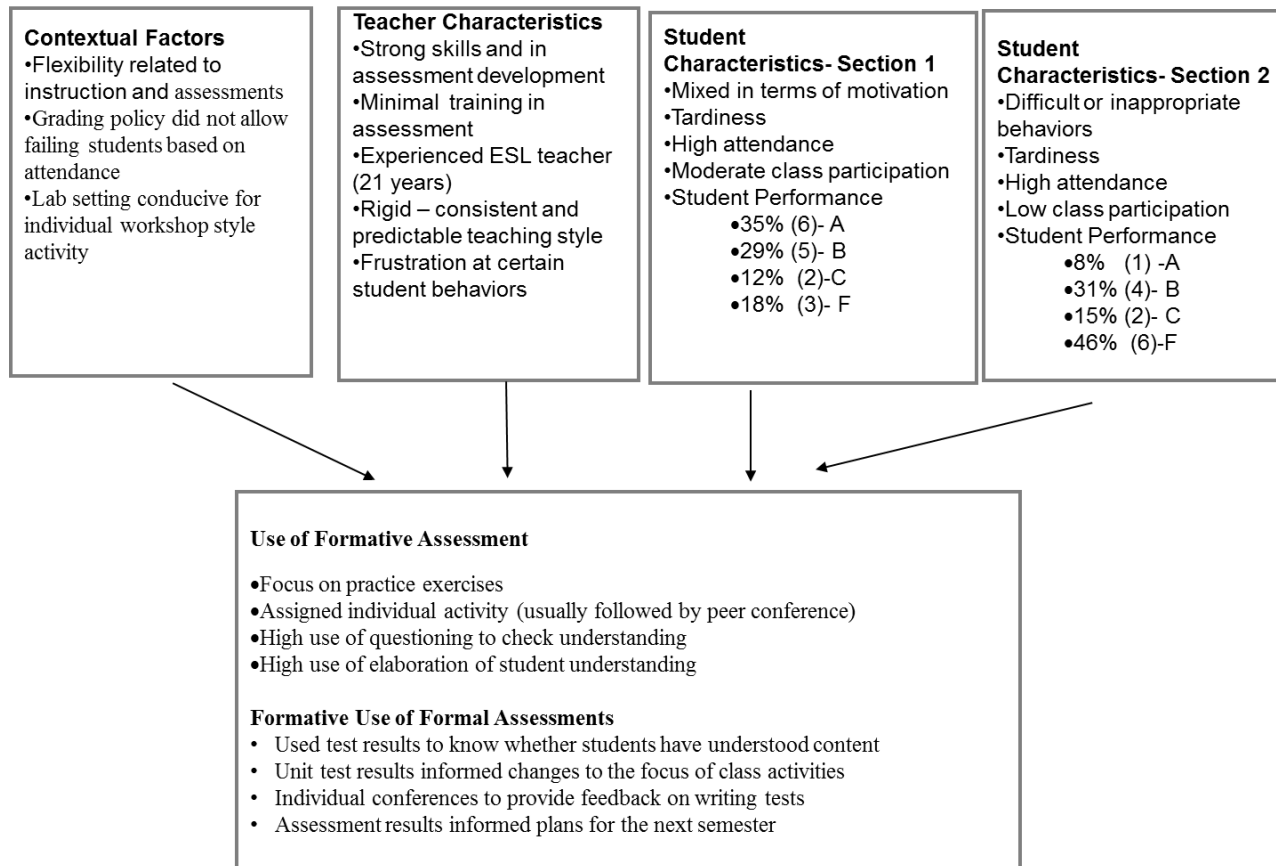


Figure 6. Formative assessment practice of teacher participant Thomas. Contextual, teacher, and student characteristics, and attributes of formative assessment pertaining to Thomas garnered from the teacher interviews and observations of sections 1 and 2.

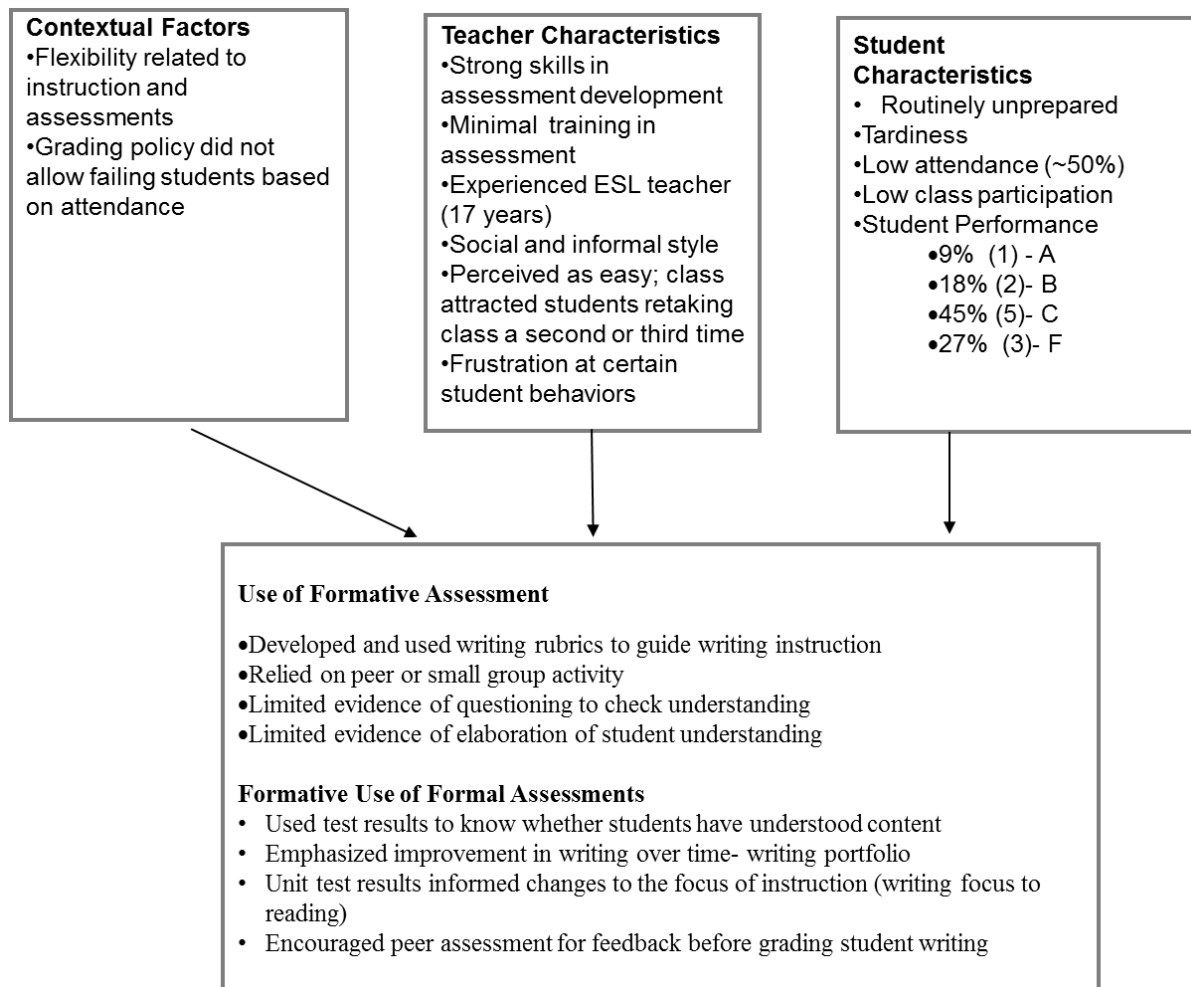


Figure 7. Formative assessment practice of teacher participant Daniel. Contextual, teacher, and student characteristics, and attributes of formative assessment pertaining to Daniel garnered from the teacher interviews and classroom observations.

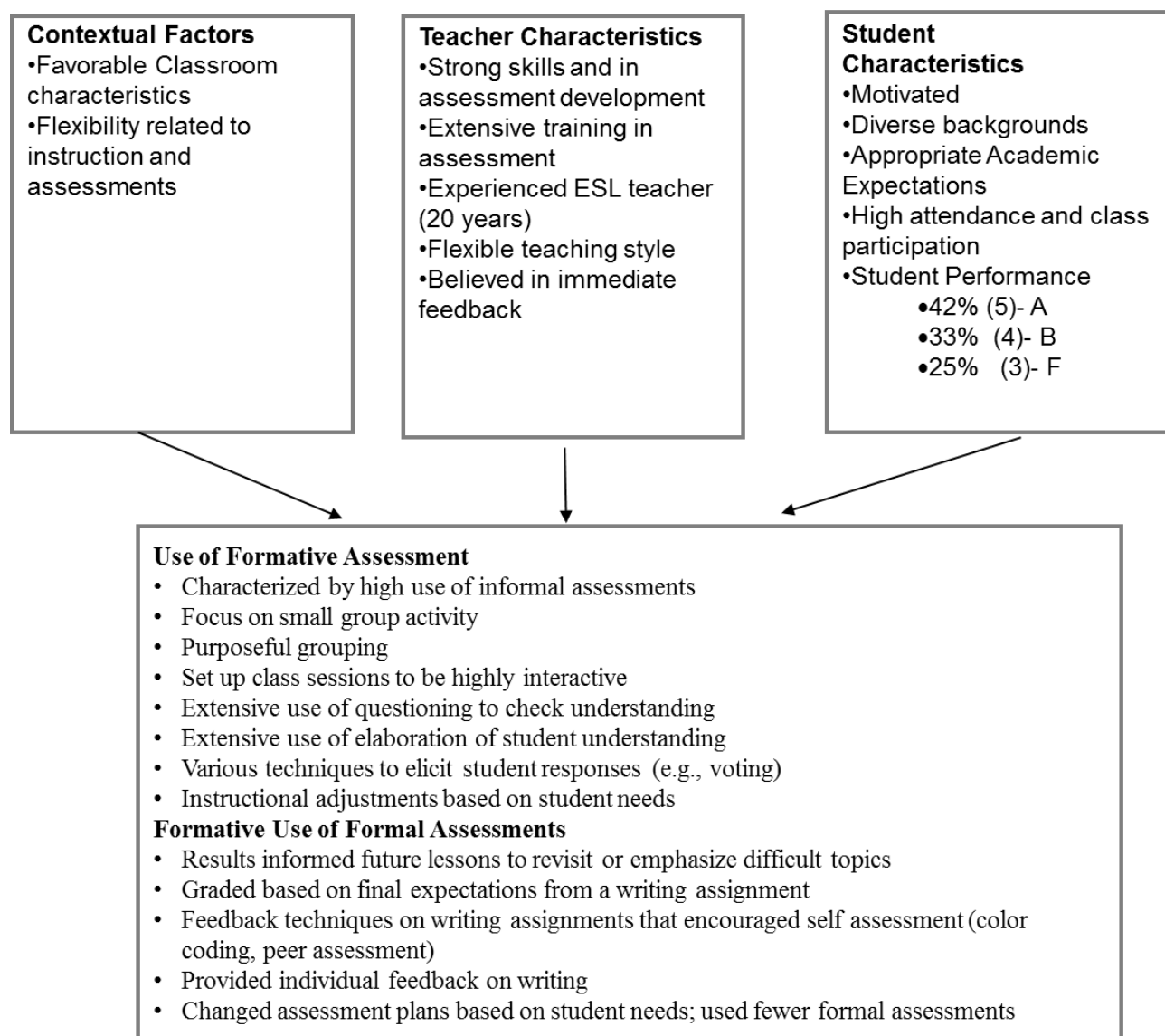


Figure 8. Formative assessment practice of teacher participant Linda. Contextual, teacher, and student characteristics, and attributes of formative assessment pertaining to Linda garnered from the teacher interviews and classroom observations.

How did the three teachers use classroom assessment to inform instruction?

Figures 6, 7, and 8 also illustrate the nature of instructional adjustments teachers made as a result of formal classroom assessments. The interview protocol included two questions that specifically address the above research question: “In your daily instructional practices, how often do you adjust instruction based on evidence that students know/don’t know the content/skills that

you are teaching?” “How often do you adjust instruction based on evidence from formal assessment that students know/don’t know the content/skills that you taught?” Participants’ responses to the former question informed a larger question that asked what their formative assessment practices were like, and how they made instructional modifications; the following section synthesized how teachers used formal classroom assessments to inform instruction.

All three teachers relied on formal assessment evidence to know whether students had learned the skills/content taught over a period of time. All of the classes included several writing assignments that were part of the writing portfolio, as well as end-of-unit tests in three sections, both of which served as evaluations of student learning (summative purpose) as well as mechanisms to elicit student understanding (formative purpose). Thomas connected patterns in student responses on tests to his instruction regularly; however, he was not able to revisit the topic often because of time constraints and the need to get through the course material. He was nonetheless able to make a change that was a direct result of examining students’ poor performance on the first two reading and skills tests. Instead of providing free writing time when the class started, he provided test-like reading exercises to help students improve their reading skills.

An interesting aspect of Thomas’ instructional adjustments was his use of the semester’s assessment information to reflect on his plan for future WC3 classes; his administrative role may have influenced his ability to think about making more substantial changes to the curriculum to better suit student needs.

Daniel made a similar change to his instruction by relying on reading/skills instruction more. He spoke to other WC 3 teachers and made the decision based on a consensus that students needed additional instruction and support for reading/skills more than they did for writing.

Daniel used assessments as a way to gauge students' level of commitment to the class; since he had many absentees, tests were the only way to know if students had learned. He conveyed a more hands-off approach to feedback on formal assessments. He expected students to engage in peer assessment to improve their drafts and hand them in to get a grade.

Linda claimed that formal assessments were an important source for informing instruction, but that she tended to rely on informal assessments much more. She used patterns in student performance to revisit gaps in understanding while covering the new unit, but not unlike Thomas, she too cited time as a constraint that kept her from re-teaching material. Formal assessments helped inform her of issues that could be covered in the next lesson as opposed to re-teaching a concept or revisiting a test.

She used students' essay drafts for the portfolio to provide individual feedback, and as opportunities to see if students had understood skills as well. For example, while she was reading student essays drafts, she noticed that students had made run-on and sentence fragment errors. Since this was established as a pattern, she revisited the related content and provided an exercise using excerpts from students' drafts. Even though she preferred informal assessments over formal ones, she also spoke about the need for grades at least every now and then, as students cared about grades.

How did the three ESL teachers use the metacognitive data they received in their formative assessment practice?

The purpose of this question was to gather information on teacher perceptions of the summary JOL data that they received for each unit test. Due to the exploratory nature of this aspect of the study, teachers were encouraged, but not expected, to make instructional adjustments based on the JOL-Performance summary data (See Appendix F for a sample

summary report prepared for Thomas's class on Unit 6 JOL and test results, section 1). During the follow-up interview, participants were asked whether they reviewed the JOL summary data, what they learned from the data, and if they used the data in any way. In short, teachers were able to review some of the metacognitive data they received, but none of them used the data in any way in their instruction. They cited time constraints that did not allow them to consider the summary data in providing feedback to students.

In their interview responses, they reflected on the information presented by the data and shared their views on it. However, due to time constraints, they were unable to thoroughly review the summary data, and hence did not use the data to directly inform their instruction. While all teachers saw discrepancies between JOL and actual scores, they did not reflect on the extent to which this may or may not have indicated a gap in students' metacognitive judgment of learning versus actual learning and performance.

Thomas commented that the data were hard to interpret at first. Following a brief session where the researcher explained the charts to him, he said that he found it interesting that the students who tended to well in his class often underestimated their skills; and he saw overconfidence in the students who did not do well.

I like to point that those who underestimated are from the better students too. They did well because they had a sense they weren't prepared and kept working. It made sense, I don't think there was any, it was nice to see the comparison between perception and preparation and actual results. It was nice too to be able to the last one you sent me, the results with the student numbers, if you said the student number, say student number 7 had high estimation but poor scores, I could go look at who the student was.

Daniel felt student judgments of learning seemed higher than expected, given the student make-up of this particular class, which was characterized by low attendance and participation, and the fact that a few of them had previously failed the class. He did think students were honest in their responses and appraisals, but that they were unprepared for difficult tests.

Linda provided feedback on the timing of the administration of the JOL. She said students struggled with providing predictions before taking the test; they reported to her that they would have liked to evaluate test difficulty after taking a test. Based on a cursory review of student responses to the questionnaire, she noticed that student JOL were lower than she expected, especially in the beginning of the semester. She added that had her other class (a typical class) taken the same questionnaire, they would have likely provided higher ratings but performed lower.

The interview protocol also included four questions to probe teachers' perspectives on their students' metacognitive awareness and skills and other sources of support for student metacognition in this setting. About their students' metacognitive skills, Thomas said students who possess high metacognitive skills were a minority, and that it is lacking in most students; Daniel said students are capable of good metacognitive skills but that fostering those skills took time, and he believed his students improved their metacognitive skills as they went on to take academic courses as undergraduates, adding it was probably true for 50-60% of them, and some students would never get there; Linda said for 9 out of 12 students, "their awareness of what they needed to learn, how they needed to learn it, what they needed to do were up there, they had it all" indicating that students who did well in class possessed the metacognitive skills needed to be successful in class.

Teacher responses indicated there were no formal lessons that targeted students' metacognitive skills. Yet, teacher comments indicated that they encouraged self-assessment and modeled learning strategies to support student metacognition. According to Thomas, teaching metacognitive skills were embedded in instruction of content and review activities.

I don't have any formal lessons or presentations that we're going to talk about this, but in preparation for a test...as we get ready for a test, I will suggest strategies. Or when we're going through an exercise, if we're going through an exercise in vocabulary, focusing on prefixes, suffixes, roots, things like that, those are all strategies so in the context of presentation, presenting a particular point, there would be some strategies. I do that regularly.

Daniel viewed teaching self-editing or self-revision skills as a metacognitive activity. He added that developing self-revision skills was part of the curriculum for WC3. After the mid-term, when the class begins to prepare the writing portfolio, they work on editing drafts where students need to examine their work and make changes. He felt that it is not possible to teach metacognitive skills, but building confidence in students about their abilities is an important goal. Towards the end of the follow-up interview, when Linda heard the question on what her perceptions were of her students' metacognitive skills, she felt that educational/ academic experience had a lot to do with metacognitive awareness and metacognitive strategies. She provided examples of strategies to enhance students' metacognitive skills that she used in class that were formative in nature. For instance, in a strategy called five-minute evaluation, she asked students to write down what they had learned and what questions they still had, as a form of self-assessment as well as an evaluation of instruction.

The course included a writing portfolio required of each student that documented students' work on essay drafts and a reflection paper on what students learned and where they needed to improve. The writing portfolio represented a substantial portion of the course grade, about 50%. Thomas reflected on the value of writing portfolios in documenting student improvement. In the follow up interview, he said that he spoke to an administrator to introduce writing portfolios to students in WC 1 and WC 2. Daniel valued writing portfolios as well, as he noted it was "absolutely essential" for students to see their improvement.

Chapter 5: Discussion

The purpose of the study was to examine the formative assessment practices of three teachers in a postsecondary ESL setting. Using a sociocultural theoretical framework, the study employed multiple data sources and multiple methods to understand the individual and contextual factors involved in their formative assessment practice. The teachers participated in two interviews during which they responded to questions related to their assessment and feedback practices, how they used classroom-based assessments to support learning and improve instruction, their beliefs about learners, and influence of and the program/department in their instruction. The researcher examined teacher-student interactions in eight classroom observations to understand the nature of formative assessment and support information gathered from the interviews.

The study also sought to explore the potential of student metacognitive judgments of learning in the process of formative assessment. To this end, students responded to a pre and post survey on their metacognitive beliefs, perceived difficulty of the English language domains, and influences on writing. Paired-sample *t*-tests were conducted to examine differences in pre and post responses. Students also completed a four-item judgment of learning (JOL) questionnaire prior to each of the three reading and writing end-of-unit assessments. Student ratings of their preparation for the test, knowledge of the material, difficulty of the test, and confidence about their performance were correlated with their performance and were provided to the teachers in the form of a narrative report that analyzed the relationships (accuracy) and provided scatterplots illustrating each students' data. Correlational analyses were conducted to examine the relationship between the JOL and test performance. At the follow-up interview,

teachers were asked about their perspectives on the reports and whether they incorporated any of the information in their feedback or discussion with students.

The following sections summarize and discuss findings presented in Chapter 4 in relation to the literature, limitations, and implications for future research.

The relationship between judgments of learning and performance

Students ratings on 'how well do you know the material taught in Unit 6/7/8?' were significantly and positively correlated with their writing and skills test scores for Unit 6, 7 and 9; even though writing tests were not directly testing the content covered in each unit, knowledge of the vocabulary, grammar, and reading skills associated with the unit may be related to student performance in writing essays in a general sense. Interestingly, knowledge, preparation, and confidence JOL were significantly related to each other; however, this did not translate into clear or consistent patterns in the JOL-test score relationship for preparation or confidence.

The results for knowledge JOL suggests two points- one, it may have been easier for students to provide relatively accurate measures of how well they knew the material using the content included in the unit as a reference. Students' knowledge judgments may also have been influenced by what they heard from teachers. Thomas and Daniel provided ample opportunity for students to learn, practice, and review the concepts in each unit in class using a combination of lecture and practice exercises. Both teachers connected class activity and course material to the unit test on multiple occasions during observations. "If you don't' (read the chapter), you will not do well in the test" (Thomas, Observation 1). Both teachers acknowledged difficulty of material and reminded students that unit tests get progressively harder and reiterated test taking strategies for students and often connected the difficulty of material to what they can expect from academic courses at the university.

I want you to practice doing the kinds of questions you will get on tests...the tests you are going to see now... are going to be difficult than the last one, because ... you need to show (teachers) ... that you are able to read and understand- because once you get into first 100 level classes ... it's going to be difficult reading and you need to understand if you want to get through these courses.

Two, the other judgments of learning, viz., preparation, difficulty, and confidence, could be more subjective and personal and hence, individual learners may have differed in their interpretation of what low or high level of preparation or confidence represented. Or, they may have tapped into previous experiences in other WC classes or in class activities related to the unit test. If all students were accurate in their JOL, i.e., if students consistently reported how well they prepared or how confident they are that they will perform well, the data would show a clear linear positive correlation. The findings of the study did not show such a clear pattern; however, as noted in the literature review in studies that examined JOL and performance, high achieving students tend to underestimate or be accurate in their judgments, and low achieving students overestimate their performance (Miller & Geraci, 2011; Nietfeld et al, 2005).

Students have the developmental ability to provide accurate judgments of learning; it is well documented that learners become progressively better at monitoring their learning accurately, but developing the tools to provide accurate judgments of learning need to be taught and can be taught (Schneider, 2008). Dinsmore & Parkinson (2013) found that students used multiple personal and task related factors in determining confidence judgments including prior knowledge, text characteristics, item characteristics, and guessing. Further, students differed in the nature and the number of factors they used to determine their learning judgments. Hattie (2013) discussed the powerful role of prior knowledge especially inaccurate prior knowledge in

overconfident learning judgments and impeding new learning. In addition, students may demonstrate overconfidence by reflecting on more surface level goals of a test like completing it on time or meeting a certain length, attending a class review of the test, etc., as indicators of learning as opposed to engaging in a reflection of more substantive aspects of learning and performance. These factors may be especially pronounced in English language learners who may rely on their educational experiences in their cultural/educational contexts or prior experience to frame their learning goals and learning strategies. Thus, inaccurate judgments of learning may be indicative of students preparing for tests with goals and learning strategies inconsistent with the goals of the class in an American higher education context.

The lack of any significant relationship between test difficulty judgments and other JOL and test scores were harder to interpret. It was expected that students would rate a test to be difficult if they were less prepared, less confident, or demonstrated low performance. Students assigned fairly moderate ratings (average) to test difficulty regardless of performance or other JOL. Student may have found it hard to judge test difficulty prior to taking the test even though they had a chance to look at the exam questions. Perhaps students judge test difficulty after taking the test, and a judgment of difficulty before may be a guess at best or based on conversations with teachers or other students about the test. Linda in the follow-up interview provided feedback on the JOL questionnaire that may provide some insight into students' thinking about test difficulty. Her students expressed they were unsure about completing the test difficulty judgment stating that they did not know whether or not the test was difficult until they had a chance to respond to and complete the test. Daniel also referred to having received a similar comment from students in his class and that a student chose to leave the difficulty JOL

blank on two occasions since he felt he was unable to respond to the item appropriately. Delayed judgment of test difficulty may have yielded different results.

The findings may also be explained by weaknesses in the technical quality of the JOL questionnaire. Factors that affect correlations including non-linear relationship between variables, presence of outliers, and the distribution of the two variables, and measurement error (Goodwin & Leech, 2006; Howell, 2012) are factors that cannot be overcome. Also, students may not have a common understanding of low and high preparation, confidence, knowledge, and test difficulty making the judgments non-equivalent and thus led to low correlations. In other words, the rating scale for all items, 10 represented “very well prepared”, “very confident”, “very difficult”, and “know very well”; students may differ in their definition of what highest level of each learning judgment represents.

Further, research on the reliability of students’ learning judgments over time or over different performance situations has been mixed (Dunlosky & Thiede, 2013). Dunlosky & Thiede argue that unstable judgments are still informative. If students’ learning judgments are unstable, examining the changes in their learning judgments can inform the process to improve their judgments’ accuracy. In this study, but for knowledge ratings, learning judgments did not show a consistent pattern with performance. The limited sample size did not make it feasible to examine the stability of ratings among individual students over time, and hence, examining the stability of confidence judgments is an area of future research (Schraw, 2009).

Finally, students provided prospective judgments of learning which may be less predictive of performance compared to retrospective or delayed judgments. JOL researchers have used both prospective, retrospective, and delayed judgments; they have found that delayed judgments (learning judgments provided after a period of time has passed since taking the test)

tend to be the most accurate (Schneider, 2008; Schraw, 2009). Since the purpose of these JOL was to explore their potential for formative assessment, prospective judgments were theoretically appropriate since they represent student notions about their learning and preparation for the test. Prospective judgments provide a measure of their metacognitive awareness based on students' planning and preparation without the influence of the experience of taking the test and having an idea of their performance.

The aim of the exploratory investigation was to assess the potential of JOL for use in formative assessment. The preliminary findings from this study highlight many issues documented in JOL research as narrated above. There is a need for research on a larger scale examining how ESL students fare in the accuracy of learning judgments and issues to consider related to timing of the learning judgments, how teachers make sense of the learning judgments, and ultimately, how they can use it in their formative assessment practice.

On account of implementation issues that are described in detail in the limitations section, teachers did not use the JOL data they received from the researcher. Thomas requested the researcher to go over the scatterplots in first unit test JOL report for both sections since he had trouble understanding the graphs. The researcher explained the charts and the narrative that Thomas said was helpful. During the follow-up interview, he said the format of the report was helpful because it helped him connect the learning judgments and performance to individual students- "It was nice too to be able to... the last (report) you sent me, the results with the student numbers, if you said the student number, say student number 7 had high estimation but poor scores, I could go look at who the student was" and interpreting the results for individual students made sense to him. Linda and Daniel reported that they did not get a chance to go over the reports in detail on account of a lack of time. Together, the findings and teacher comments

suggest that timing of administration, format of the report, and time constraints related to using a JOL tool in an assessment task are important considerations for future research on JOL application in formative assessment.

Further, the interpretations made by the teacher participants about inaccuracies in their students' learning judgments and performance highlight an important issue - how teachers make sense of students' learning judgment inaccuracies. These interpretations may have been peculiar to the setting, but highlight an important issue in using learning judgment measures in classroom assessment. Consistent with the literature, Linda, whose students were mostly high performers, noted that student predictions or judgments tended to be lower than their performance. Thomas viewed underestimation on the part of high achievers as desirable so that the student may expend more effort in preparing and performing well in the test. At the same time, Thomas and Daniel attributed the tendency to overestimate learning judgments to arrogance rather than genuine weakness in students' awareness of their learning- "they thought they knew it all" (Thomas-Interview 2). Just as students have multiple factors to consider in providing their learning judgments, teachers may have multiple factors based on which they interpret students' learning judgments. Attributing overconfidence to personal attributes of students with learning deficits may further impede their learning by discouraging teachers to invest in providing appropriate feedback to address students' inaccurate learning judgments. Therefore, the role of teachers' perception of students' learning judgments is an important consideration in the application of learning judgments in formative assessment.

No Change in Students' Metacognitive Beliefs Related to Writing

The second finding of the study was that there were no significant changes in students' metacognitive beliefs about writing over the course of the semester. The purpose of the survey

was to examine whether students' metacognitive beliefs improved over the duration of the course. The course marks a leap from basic and intermediate competency in writing to a level of proficiency comparable to university students in freshman level writing courses; students engage in intensive writing during the semester, and the course may be viewed as an intervention in itself to enhance their skills in writing. Students' mean scores on the metacognitive beliefs scale showed an increase in the positive direction, but did not reach significance. The results may indicate two issues- one, in a span of two months (approximately), students did not experience a significant change in their beliefs about their writing skills; this may be on account of stability in these beliefs that may need a targeted intervention over a lengthy period of time to bring about any significant change. As such, results may indicate that a significant change in writing metacognitive beliefs may not be a natural outcome of engaging in a semester of WC3. Changing students' metacognitive beliefs may require evoking their current beliefs, encouraging reflection of the same throughout the semester, and purposefully communicating to students their skills in different aspects of writing as they achieve learning targets making them conscious of the changes. Two, a small sample size may have affected the results; for most statistical analyses that compare means, a sample size of at least 30 is recommended to produce meaningful results. Thus, the findings must be interpreted bearing in mind this limitation.

In addition to metacognitive beliefs, students did not indicate a significant change in their perceived difficulty of reading and writing domains of the English language. The limitations of sample size notwithstanding, student responses indicated their perception of reading and writing difficulty did not change as a result of participating in WC 3. The reading and writing content is substantially different from the content covered in WC 1 and 2; in Daniel's words: "In WC1 they learn how to put together sentences, and make paragraphs, and if they started below that, they

learned grammar, how to make the simplest of sentences and form the letters. So a lot of what we ask them to do in WC3 is brand new and scary”. WC 3 content is closer to undergraduate level writing courses that students may have found especially daunting and challenging; this may have affected their perception of difficulty of reading and writing; they may have come in with the knowledge and beliefs about their reading and writing based on their experience in previous classes; the post survey results may have changed the definition of difficult and hence, even though students may have gained competence in reading and writing, their metacognitive beliefs and perceived difficulty about reading and writing may have changed to reflect the newer expectations of WC 3.

The Nature of Teachers’ Engagement in Formative Assessment

Inquiry into the teachers’ formative assessment practice was framed using sociocultural theory to uncover the roles of different elements like the ELP policy, teachers, and students in the process of formative assessment. All participants demonstrated various aspects of formative assessment embedded in their daily instruction. Each teacher’s practice and its success was shaped by their beliefs and teaching style and student attributes that facilitated or hindered their practice. These findings point to the value of investigating formative assessment and its impact on learning by using a holistic approach where individual, contextual, and cultural aspects can be better understood (Schoen, 2011). Sociocultural theory is advantageous to understanding the role of student characteristics like their academic expectations and cultural educational experiences in a learning context (Elwood, 2007); sociocultural theory also supports the use of multiple data sources and methods to understand a phenomenon (Schoen, 2011). In a setting where teacher and contextual expectations of students may be different from students’ background experiences and expectations, investigations of formative assessment stand to gain

rich information on these differences that may not become apparent in traditional, quantitative studies.

Findings are consistent with other similar inquiries based on sociocultural theory that found teachers implemented a variety of formative assessment attributes (Crossouard, 2011; Pryor and Crossouard, 2010; Sardareh & Saab, 2012; Willis, 2012). In each study, like this one, the nature of classroom interactions, garnered mostly through observations, arrived at the same conclusion that it takes both teachers and students to successfully implement formative assessment; in each study, the researchers interpreted the role of the teachers in fostering dialogue as paramount; for example, Willis (2012) asserted that teachers needed enable student participation and provide learners a sense of belonging for learners in the formative assessment process. However, the studies only recognized the active role of students in the learning process. Thomas, Daniel, and Linda showed differences in their interactional style and how they negotiated the role of students in class. Each teacher also emphasized the attributes that students brought to class as important determinants in the learning process. In this study, teachers brought students' attributes to bear on the success of formative assessment interactions- attributes they did not necessarily perceive to have control over.

Linda was a highly experienced teacher with extensive training in assessment and strong skills in developing and using assessments. She was explicit in her use of FA principles, and she showed the most variety in terms of techniques and use of informal assessment to inform instruction; her practice was guided by seeking evidence of student learning and errors in their writing and reading assignments; she made several instructional adjustments based on that information and followed a pace guided by student learning. She fostered dialogic interactions in reviewing practice exercises where students learned from each other as she led class sessions;

she appreciated that students learned from each other and that contributed to the success of her ability to implement her techniques successfully. She navigated classroom interactions through her role as an assessor and an authority of the content at the same time she was a learner, gauging student understanding and planning the direction of instruction, “an active modifier” (Poehner, 2011). In her class, students were “insiders” who were central to the process of formative assessment (Willis, 2012). She added though this success did not typically occur in her classes “but when it does, it is just golden”.

Linda also had other facilitating factors in her favor. The program provided her the flexibility to make changes to the assessment structure. She chose course materials that matched her teaching style, and implemented a flexible assessment and grading plan while ensuring that she was meeting the curriculum and course requirements. The classroom physical characteristics, the writing wall, was conducive to her method of demonstrating concepts visually as she took the class through a continuous cycle of questioning, gathering student response, acknowledging and elaborating on that response before proceeding. Compared to the other classes, there was a noticeable difference in the movement and conversation that occurred in this class as students were drawn to participate and contribute to the learning process.

Thomas and Daniel engaged in a similar cycle of questioning, gathering student response, and elaborating on that response to reiterate the learning concept during lecture and while reviewing practice exercises. Unlike Linda though, their interactions with students were markedly different. Thomas introduced more reading exercises upon learning from students’ poor performance on the reading tests; he believed they needed more practice in reading and made the instructional decision to provide more learning opportunities; but, he placed the onus of using the opportunity on the students. His instructional adjustments reflected his rigid style and

were characterized by modifications that were predictable and consistent in contrast to Linda's flexible, evolving style.

Daniel was similar in his approach where he provided instructional situations to facilitate learning, but it was up to students to take action that led to learning - "I treat them like college students. Here's the homework, I suggest (students) do it." (Daniel, Interview1). His preference for peer assessment as a source of feedback to students reflected this style. However, like Thomas, he experienced frustration on account of student absences and non-participation, factors that negatively affect the value of peer assessment as a learning activity. He felt students thought he was an easy teacher, a remedial teacher who attracted students who were retaking the class in hopes of a positive outcome. Although he made instructional adjustments based on assessment results, there were no indicators that Daniel made instructional decisions targeting the particular aspect of dealing with students who have previously demonstrated failure in the same content.

The role of student attributes in formative assessment. This study's findings suggest the possibility of a mediating relationship between student attributes, formative assessment, and student outcomes, especially in higher education/postsecondary settings where students are expected to take ownership of their learning; attributes that support or hinder student outcomes seem to play a role in the teachers' implementation of formative assessment that in turn may magnify or diminish student outcomes. Linda's implementation of formative assessment was supported by desirable student attributes; at the same time, her student-focused instructional style served as an advantageous learning environment for students.

Teachers attributed student success and failure to various individual characteristics including effort, participation, and academic skills. None of the teachers attributed failure to a lack of ability or genuine learning difficulties. They had a shared understanding of these

characteristics that at times seeped into their instructional and assessment practice. Thomas and Daniel found themselves making changes to planned activities and experienced frustration when students' negative attributes manifested as low class participation or students failing to do class work and homework. Individual student attributes collectively became class characteristics that were powerful enough to make or break the instructional routine. For example, Thomas described dismissing a class early on account of many students having come unprepared to class; their behavior, specifically students in Section 2, was disruptive and demotivating for him and the other students to the extent that he said there were times he did not look forward to class sessions. Linda voiced similar concerns about another 'typical' class she was teaching, but in contrast, Linda's descriptions of her formative assessment and feedback practices were example after example of the motivation and effort of the majority of her students who have the intent to learn.

Linda believed that the class dynamics helped keep up an environment of positivity and success. For the students who failed the class, she explained there are layers of reasons that cause failure - a lack of understanding of academics and expectations in higher education, a government sponsoring their education, and the convergence of being young and coming from a restrictive culture to a less restrictive culture and an educational setting with lenient policy. Perhaps, it was the preponderance of similar student characteristics that made Daniel's and Thomas' classes difficult.

The central role of informal assessments in learning & instruction. Findings indicate that assessments were an integral part of learning and instruction in this setting. All three teacher participants used a variety of assessment tools in the form of worksheets, practice exercises, homework activities. Teacher-student and student interactions were defined by activities that

included working on some form of an assessment tool that was typically not graded. In fact, the grading scale used in this setting mirrored the assessment policy and grading scale documented in several studies in the ESL literature. For example, the emphasis on process oriented approach, use of portfolios, and relatively low stakes on formal tests are consistent with other ESL environments studied in the literature (Chen et al, 2007; Cumming, 2009; Davison & Leung, 2009)

Brookhart (2013) and Supovitz (2012) referred to the types of assessments used in this setting as short cycle assessments- assessments that are teacher-made, low-stakes, and closely aligned with instruction. The three teachers embedded these assessments into daily instruction, and involved sharing criteria for success, providing immediate feedback and student participation, and opportunities for student self-evaluation, thus providing the perfect ground for teachers to implement a high level of formative assessment (McMillan, 2010). All teachers described they saw improvement and success in students who attended class and completed classwork and homework; perhaps these students benefited from the many opportunities to practice and review learning. Conversely, students who failed were predominantly those who missed classes and hence missed the learning opportunities.

Hattie & Timperley's (2007) model of feedback to reduce the discrepancy between a student's current understanding and desired goal was apparent in classroom interactions. All three teachers engaged in review activity and facilitated learning through questioning and elaborating on student responses. They also demonstrated the four levels of feedback posited by Hattie & Timperley (2007). They most commonly provided both task-level and process-level feedback by checking what students knew and elaborating on students' responses to reiterate the process underlying the concept, both desirable types of feedback. There were instances of

positive feedback but in no instance was it geared toward personal attributes of students, and there were no instances of negative feedback- both undesirable forms of feedback.

Teachers' informal and formal feedback practices illustrated several themes that have been documented in the literature on ESL feedback practices. Daniel and Thomas described their strategy was to avoid corrective feedback on students' writing instead providing deliberately vague feedback as they wanted students to figure out next steps or seek help from the teacher; Linda said her feedback comprised of questions to students about their writing, although she added she corrected student errors too. Studies in postsecondary ESL settings lend support to Thomas and Daniel's rationale. Providing corrective feedback may not be helpful to students because revised drafts based on such feedback do not necessarily reflect student learning but teacher's revisions (Ashwell, 2000).

Studies have also indicated that teachers do not know how students perceive, understand, or utilize feedback (Bailey & Garner, 2011; Ferris et al, 1997). There was some indication this was true for the teachers in the study. The literature has documented that ESL students have reported that they preferred clear feedback on grammar, vocabulary, and mechanics (Hedgecock & Lefkowitz, 1994) or task-based, corrective feedback (Hattie & Timperley, 2007). They struggled with implicit and vague feedback, both verbal and written, and failed to understand or use feedback the form of questions or vague comments as requiring action on their part (Brice, 1995; Higgins et al, 2002; Panova & Lyster, 2002).). This is in contrast to Daniel and Thomas' feedback style. The study did not adequately address cultural differences or differences in academic expectations of ESL students and teachers that could affect how, whether and to what extent students accept feedback and act on feedback. This is a critical element to examining the

success of formative assessment that needs to be examined in future ESL based formative assessment studies.

Limitations

There are several limitations in the study design that affect the nature and extent of the validity and generalizability of the above findings. The study examined the practice of formative assessment by investigating in-depth a sample of three teachers in one learning context. The nature of the context and its participants were studied using a case study design with qualitative and quantitative methods. Such an investigation affords great strength, validity, and reliability to the findings but severely limits the extent to which they may be generalized to even a similar learning context.

Study Design and Implementation Related Issues

The mixed method case study design allowed for in depth investigation of the three teacher participants and their students. A major drawback with case study design is the inability to generalize findings to a larger context. Findings are based on evidence from only three teachers in a single ESL context; as such, the study may not be generalized to a similar population, and it was not the goal of the investigation to do so. The aim of the study was to inform theory and the nature of mixed method case study design made it an appropriate design for that (Yin, 2014).

The design assumed a high participation rate from students in order to adequately address the quantitative research questions. Response rate for the pre and post surveys was low. The JOL questionnaire participation was better, still because all participating students did not take all the JOLQs and all the sections did not share the common assessments, the analyses were conducted separately for that section which severely limited the scope of the analysis to address the

research questions. This limitation also affected the researcher's ability to examine mixed methods research questions related to change in students' metacognitive beliefs in relation to teachers' use of metacognitive judgments in their formative assessment practice.

The researcher selected the study context purposefully to address the need for formative assessment research in this context; the context was also convenient, practical, and accessible to the researcher to study. The researcher chose the teachers based on their expressed interest and willingness to participate in the study. Their willingness may represent specific attitudes and beliefs that may have influenced the study findings. Similarly, the student participants represent a distinct demographic of English language learners with a majority of students representing one country, Saudi Arabia. As such, findings may be indicative of traits and behaviors unique to this demographic of students. Students' cultural and academic background was an important student attribute that teachers described in the interviews. The study did not sufficiently examine the similarities, differences, and the interplay between different cultural expectations guiding the teachers' and students' behaviors in a learning context aligned with the teachers' culture.

Finally, using English to communicate with students and collect data about complex psychological beliefs poses questions about the validity of students' responses given their limited proficiency in the language. The researcher made every effort to provide students opportunity to ask questions about the study; she explained the purpose of the study, their rights and scope of participation in addition to providing information in writing and following all ethical expectations set by the IRB for human subjects research.

Differences in the nature of the three classes restricted the use of all student responses on the JOLQ in the quantitative analysis. Linda's class used assessments that were different from the other two sections; as such, these data were examined separately; this reduced the sample

size and consequently the range of statistical analyses available to analyze the relationship between the JOLQ, pre and post survey, and student performance. The response rate to the online surveys were approximately 35%; given the total number of students was 53, there were only 19 - 21 usable responses and even fewer for both the pre and post surveys. The researcher took several measures to increase participation rate as recommended in the literature (Dillman, 2000) including sending personalized invitations to the survey, verbal announcements and reminders by the teachers, providing incentives for participation. However, the measures did not significantly increase the number of student participants taking the pre and post survey. Therefore, findings related to the quantitative measures can only be considered preliminary and exploratory.

Implementation Issues. There were unanticipated implementation related issues that affected the data collection and analysis. First, the study was implemented three weeks after the academic semester started; as such, three classes had taken one unit test that was also supposed to be included in the study. On account of the delay, only three unit tests were included in the study of JOL performance data. Second, the researcher expected to provide JOL data to teachers within 24 – 48 hours after students took the tests to enable teachers to provide feedback related to JOL to students if they desired. Although Daniel and Thomas administered the unit tests on the same days, they did not follow a set timeline for grading and feedback. This especially an issue with Linda who did not follow a set schedule for administering unit test equivalents. Also, some students missed the test on test day and took make up tests between 1 – 7 days after the test date which further delayed the researcher from sharing the summary reports with teachers. The interview protocol did not adequately cover teacher experiences with the JOLQ. For example, the teachers were asked whether they viewed the JOLQ results and if they used it to provide feedback. Including probing questions to get at why teachers did not use the results may have

resulted in gaining more insight into teachers' perceptions of the JOL reports and their use in formative assessment.

Analysis and Interpretation of Findings

The researcher employed quantitative analysis procedures based on their appropriateness to address the research questions. However, they represent one of many ways to examine and interpret relationships between variables. For example, examining correlational patterns of learning judgments and performance separately for high and low achievers may have provided fine grained information on the accuracy of learners' judgments. Similarly, examining the stability of learning judgments over time using a repeated measures Analysis of Variance is a key component to understand how learners' judgments vary over time and over different tests.

As for qualitative analysis procedures, the researcher chose a coding and analysis procedure by referencing guidelines provided by qualitative researchers (e.g., Krefting, 1991; Bogdan & Biklen, 2007). The procedures followed in this study reflect a combination of several approaches to qualitative data analysis that would help adequately address study goals and research questions. The researcher also considered her comfort level and previous experience with qualitative data in identifying a suitable method. As such, another approach may have led to different interpretation of data and presentation of findings.

The role of the researcher. Findings represent the analysis and interpretation of the researcher whose perspective and viewpoint may be different from another researcher conducting a similar investigation. The researcher also collected the data that informed the analysis and interpretation. While it is not possible to measure or eliminate the impact of the researcher's role in the findings, several steps were taken to minimize personal bias and broaden

the researcher's perspectives and enhance the dependability of findings (Krefting, 1991). These steps enhance the trustworthiness of findings.

The researcher relied on the research literature to inform all aspects of the investigation including framing the research questions, instrument development, data analysis procedures, interpretation, and presentation of findings. Pilot testing the instruments with a set of participants similar to the main study participants helped revise instruments to make them relevant and informative. Using a common interview and observation protocol allowed the researcher to bolster teacher comments with actual samples of behavior; teacher comments also strengthened the interpretation of a sample of observed behavior. The researcher followed recommendations and standards provided by reputable quantitative and qualitative methods researchers (e.g., Bogdan & Biklen, 2007; Tabachnek & Fidell, 2007) The researcher audio recorded all observations and interviews to enable detailed note taking and transcriptions. The researcher maintained a data diary to document quantitative analysis procedures and memos to record thoughts, questions, and concerns related to qualitative data analysis. The researcher also thoroughly documented and presented the analysis procedures to maintain transparency.

Implications of Findings for Research

The relationship between metacognitive judgments and student performance was not strongly established on account of the nature and size of the student sample. Teachers discerned inaccuracies in student judgments where low performers overestimated their learning, and high achievers under estimated their learning. While this is consistent with metacognitive judgments research, the teachers went further to attribute students' inaccuracies to traits related to their high or low achieving status. For example, Thomas viewed underestimation on the part of high achievers as desirable and motivating student effort, and attributed overestimation to arrogance

rather than a deficiency in metacognitive awareness. In addition to identifying, validating, and examining JOL measures for use in formative assessment, future research on the use of metacognitive judgments in the classroom must consider the role of teachers' attributions in whether and to what extent teachers may use students' inaccurate judgments to inform instruction.

Linda recognized that her students were successful because they had the necessary motivation and skills. She believed they came into class with the desired level of motivation. She described a sense of helplessness in not being able to motivate struggling students but also at the lack of buy-in from students –“teaching them (metacognitive skills or learning strategies) and then having them buy in. There's a buy in and that's where, I think sometimes there's a lack of buy in.” Teachers' views on their ability to influence college students' motivation and self-regulated learning is another area that needs to be investigated. Specifically, the potential of formative feedback interactions to enhance a student's awareness of the gap in his/her understanding of current and desired performance to improve students' metacognitive skills need to be examined.

The findings on the role of student attributes in teachers' assessment practice have implications for research on formative assessment in postsecondary ESL settings. Even as learning environments have progressed to a social constructivist view that accepts students as active and central to teaching and learning, the role of students in the formative assessment process is not well understood (Shepard, 2000). Understanding contextual and individual factors that affect teachers' formative assessment practice is valuable to how we conceptualize studies of formative assessment. Student attributes that affect student outcomes also affect student

participation in formative assessment activity thus potentially mediating or moderating the effect of formative assessment on student outcomes.

Finally, the findings of this study contribute to the growing body of literature on formative assessment based on sociocultural theory, specifically in postsecondary ESL settings. Future research should focus on larger investigations of formative assessment in ESL learning contexts that gather contextual and individual characteristics using multiple methods and sources. There is a need for a knowledge base on students' attitudes and perceptions toward assessment, for example, student perceptions toward assessment and feedback, to define students' role in formative assessment. Also important is investigating the role of ESL students' cultural and academic experiences that seem to influence their learning and performance in a native English language culture and educational context such as the United States. More research is needed to consolidate established findings related to ESL feedback practices by using the formative assessment framework.

Teachers in the study did not use the student metacognitive judgments of learning information to inform their feedback. They cited a lack of time as the reason they were unable to review or use the reports thoroughly. Thomas indicated he had difficulty understanding the report at first. The next step to investigating the use of metacognitive judgments of learning in formative assessment is considering factors that influence whether and to what extent teachers make sense of the information and use it in their feedback to students or instructional adjustments. Future research should consider the need for appropriate teacher training and expectations related to using judgments of learning in formative assessment. Such research could focus on developing a training program that introduces teachers to the concept of metacognition, judgments of learning, and handling accuracies and inaccuracies of students'

metacognitive judgments of learning. Future research should focus on the potential to embed metacognitive judgments of learning in formative assessment situations in ways that minimize the time required to administer, interpret, and use results in formative feedback. Such research may also result in tools or processes related to the use of technology (e.g., clickers) to create interactive assessment opportunities in the classroom.

Implications of Findings for Practice

Findings suggest several implications for practice. First, with regard to using metacognitive judgments of learning in formative assessment, teachers might need clear suggestions related to the use of JOL tools that are necessary to facilitate teacher use of JOL like tools in formative assessment. It may also be useful to include some form of training or orientation on what judgments of learning represent in relation to their monitoring and awareness and use of appropriate learning strategies. There is also a need to orient teachers away from attributing monitoring judgments to personal characteristics of students. Such attributions may hinder the use of JOL information formatively, but may also strengthen negative perceptions toward students that may in turn affect feedback and support to struggling students. Teachers may also need training and practice in reading and interpreting summary reports that provide individual students' and aggregate results for the class.

Findings also indicate that ESL learning contexts can benefit from formative assessment theory. Teachers in the study used a variety of techniques to support student learning and make instructional adjustments. However, students may not be aware of the process and purpose underlying the use of many of the techniques. In higher education learning contexts where students are expected to take more ownership of their learning, framing the use of these techniques with a formative assessment lens can help teachers and students focus more attention

to the purpose of the techniques. Making the formative assessment process explicit can bring about a change in the learning culture in ways that support more student participation and facilitate students to take ownership of learning.

Conclusion

In light of the findings and support from the literature, the following conclusions are warranted. With regard to the use of metacognitive judgments of learning in formative assessment, the findings showed a mixed relationship between student JOL and performance. There were significant, positive correlations between students' knowledge judgments, i.e., how well they rated they knew the material being tested, and test scores; there was no clear or consistent relationship between students' JOL regarding preparation for the test, difficulty of the test, and confidence about performance in the test. Teachers' comments regarding students' learning judgments indicated they may have preconceived notions about students' JOL. Future investigations should systematically examine how students' JOL can be informative to teachers in a formative assessment situation; further, teachers may need clear expectations and some form of training or orientation related to using learning judgments as part of formative assessment; researchers should consider teachers' attributions regarding students' learning judgments can affect the utility of JOL to inform formative assessment practice.

Teachers in this ESL setting engaged in a variety of formative assessment techniques. Short cycle informal assessments were central to teaching and learning in this context. The implementation and success of their techniques seemed to be affected by student attributes; teachers described students' irregular attendance or low attendance, tardiness and low participation, and a lack of appropriate previous academic experiences affected their formative assessment practice negatively. Teachers described students who were regular and did class

work and homework, demonstrated motivation and engagement, and had appropriate academic expectations positively affected their formative assessment practice. Teachers associated the same attributes to contrast successful and failing or struggling students. This suggests that student characteristics that affect student performance could perhaps also affect a teacher's use of formative assessment to support learning and inform instruction. Findings point to the value of sociocultural theoretical lens to inform the theory and study of formative assessment. Future research should continue to expand our understanding on the role of students in formative assessment, especially in postsecondary ESL settings.

References

References

- Abrams, L. M. (2007). Implications of high-stakes testing for the use of formative classroom assessment. In J.H. McMillan (Ed.), *Formative classroom assessment: Theory and practice* (pp. 79 - 98). New York, NY: Teachers College Press.
- Abrams, L. M., Varier, D., & McMillan, J.H. (2013, April). Teachers' use of benchmark assessment data to inform instruction and promote learning. *Paper presented at the American Educational Research Association Annual Meeting, San Francisco, CA.*
- Aljaafreh, A., & Lantolf, J.P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *The Modern Language Journal*, 78(4), 465 – 483.
- Alexander, P.A. (2008). Why this and why now? Introduction to the special issue on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review*, 20, 369 – 372.
- Andrade, H. L. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. Andrade & G. J. Cizek (Eds). *Handbook of Formative Assessment* (pp. 90-105). New York: Routledge.
- Andrade, H. (2013). Classroom assessment in the context of learning theory and research. In J.H. McMillan (Ed.), *SAGE Handbook of research on classroom assessment* (pp. 17 - 34). Thousand Oaks, CA: SAGE

- Ashwell, T. (2000). Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing*, 9(3), 227 – 257.
- ATLAS.ti GmbH. (2015). ATLAS.ti (Qualitative Data Analysis Software). Berlin: ATLAS.ti Scientific Software Development GmbH.
- Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. G.R. Hancock, R.O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences*, (pp. 93-114). New York, NY: Routledge
- Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice-Hall Inc.
- Bailey, R., & Garner, M. (2011). Is the feedback in higher education assessment worth the paper it is written on? Teachers' reflections on their practice. *Teaching in Higher Education*, 15(2), 187 – 198.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy and Practice*, 18(1), 5 – 25.
- Birjandi, P., & Tamjid, N.H. (2012). The role of self-, peer and teacher assessment in promoting Iranian EFL learners' writing performance. *Assessment & Evaluation in Higher Education*, 37(5), 513 – 533.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education*, 5(1), 7-74.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation & Accountability*, 21(1), 5-31. doi: 10.1007/s11092-008-9068-5

- Bogdan, R.C., & Biklen, S.K. (2007). *Qualitative Research for Education: An Introduction to Theories and Methods*. Boston, MA: Allyn & Bacon
- Brice, C. (1995). ESL writers' reactions to teacher commentary: A case study. *Paper presented at the annual meeting of English to Speakers of Other Languages*, Long Beach, CA.
- Briggs, D. C., Ruiz-Primo, M. A., Furtak, E., Shepard, L., & Yin, Y. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. *Educational Measurement: Issues and Practice*, 31(4), 13-17.
- Brookhart, S. (2007). Expanding views about formative classroom assessment: A review of literature. In J.H. McMillan (Ed.), *Formative classroom assessment: Theory and practice* (pp. 43 - 62). New York, NY: Teachers College Press.
- Brookhart, S. (2013). Classroom assessment in the context of motivation theory and research. In J.H. McMillan (Ed.), *SAGE Handbook of research on classroom assessment* (pp. 35 - 54). Thousand Oaks, CA: SAGE
- Brown, G. T. (2011). Self-regulation of assessment beliefs and attitudes: A review of the Students' Conceptions of Assessment inventory. *Educational Psychology*, 31(6), 731-748.
- Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education*, 18, 529 – 549.
- Boulet, M., Simard, G., De Melo, D. (1990). Formative evaluation effects on learning music. *Journal of Educational Research*, 84(2), 119 – 125.
- Butler, D.L., & Winne, P.H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245 – 281.

- Chen, Q., Kettle, M., Klenowski, V., & May, L. (2013a). Interpretations of formative assessment in the teaching of English at two Chinese universities: A sociocultural perspective. *Assessment & Evaluation in Higher Education*, 38(7), 831 – 846.
- Chen, Q., May, L., Klenowski, V., & Kettle, M. (2013b). The enactment of FA in EL classrooms in two Chinese universities: teacher and student responses. *Assessment in Education: Principles, Policy & Practice*, DOI: 10.1080/0969594X.2013.790308
- Cheng, L., Rogers, W.T. & Wang, X. (2007). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment & Evaluation in Higher Education*, 33(1), 9 – 32. DOI: 10.1080/02602930601122555
- Cizek, G. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. Andrade & G. J. Cizek (Eds). *Handbook of Formative Assessment* (p. 3 - 17). New York: Routledge.
- Clark, C.T., & Clark, I. (2008). Exploring and exposing a gap in L2 research: How socio-linguistic roles and relationships facilitate or frustrate second language acquisition. *Atlantis*, 30(1), 101- 113.
- Cleary, T. J., & Zimmerman, B. J. (2004). Self-regulation empowerment program: A school-based program to enhance self-regulated and self-motivated cycles of student learning. *Psychology in the Schools*, 41(5), 537-550.
- Crain, W. (2000). Theories of development: Concepts and applications (4th ed.). Upper Saddle River, NJ: Prentice Hall.
- Creswell, J. W., & Plano Clark, V. L. (2011). Designing and Conducting Mixed Methods Research. *Thousand Oaks, CA*: SAGE Publications

- Crossuard, B. (2011). Using formative assessment to support complex learning in conditions of social adversity. *Assessment in Education: Principles, Policy, and Practice*, 18(1), 59 – 72.
- Cumming, A., Kantor, R., & Powers, D.E. (2002). Decision making while rating ESL/EFL writing tasks: A descriptive framework. *The Modern Language Journal*, 86(1), 67 – 96.
- Cumming, A. (2009). What needs to be developed to facilitate classroom assessment. *TESOL Quarterly*, 43(3), 515 – 519.
- Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393 – 415.
- Desoete, A. (2008). Multi-method assessment of metacognitive skills in elementary school children: how you test is what you get. *Metacognition Learning*, 3, 189 - 206
- Dillman, D.A. (2000). Mail and internet surveys: The tailored-design method. New York: Wiley.
- Dinsmore, D.L., Alexander, P.A., & Loughlin, A.M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review*, 20, 391 – 409. DOI: 10.1007/s10648-008-9083-6
- Dinsmore, D.L., & Parkinson, M.M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction*, 24, 4–14.
- Dunlosky, J., & Thiede, K.W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction*, 24(4), 58-61.

- Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment Research & Evaluation*, 14(7), 1-11.
- Edwards, A., & Daniels, H. (2006). Using sociocultural and activity theory in educational research. *Educational Review*, 56(2), 107-111.
- Elwood, J. (2007). Formative assessment: Possibilities, boundaries and limitations. *Assessment in Education: Principles, Policy & Practice*, 13(2), 215-232.
DOI:10.1080/09695940600708653
- Ferris, D.A., Pezone, S., Tade, C.R., & Tinti, S. (1997). Teacher commentary on student writing: Descriptions & implications, *Journal of Second Language Writing*, 6(2) 155 – 182.
- Flavell, J.H. (1979). Metacognition and cognitive monitoring. *American Psychologist*, 34(10), 906 – 911.
- García, T., & Pintrich, P. R. (1995). Assessing Students' Motivation and Learning Strategies: The Motivated Strategies for Learning Questionnaire.
- Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning & Teaching in Higher Education*, 1, 3 – 31.
- Good, R. (2011). Formative use of assessment information: It's a process, so let's say what we mean. *Practical Assessment, Research & Evaluation*, 16(3), Available online: <http://pareonline.net/getvn.asp?v=16&n=3>.
- Goodwin, L. D., & Leech, N. L. (2006). Understanding correlation: Factors that affect the size of *r*. *The Journal of Experimental Education*, 74(3), 249-266.

- Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist*, 45(4), 203-209.
- Guenette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback in writing. *Journal of Second Language Writing*, 16, 40 – 53.
- Guskey, T.R. (2005). Formative classroom assessment and Benjamin S. Bloom: Theory, Research, and Implications. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.
- Hamp-Lyons, L. (1994). Interweaving assessment and instruction in college ESL writing classes. *College ESL*, 4(1), 43 – 55.
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42(2), 377-381.
- Hattie, J. (2013). Calibration and confidence: Where to next? *Learning and Instruction*, 24, 62 – 66.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <http://dx.doi.org/10.3102/003465430298487>.
- Hedgcock, J., & Lefkowitz, N. (1994). Feedback on feedback: Assessing learner receptivity to teacher response in L2 composing. *Journal of Second Language Writing*, 3(2), 141 - 163
- Hernandez, R. (2012). Does continuous assessment in higher education support learning. *Higher Education*, 64, 489 – 502.

- Higgins, R., Hartley, P., & Skleton, A. (2002). The conscientious consumer: Reconsidering the role of assessment feedback in student learning. *Studies in Higher Education*, 27(1), 53 – 64.
- Hirschfeld, G. H., & Brown, G. T. (2009). Students' conceptions of assessment. *European Journal of Psychological Assessment*, 25(1), 30-38.
- Howell, D. (2012). Statistical methods for psychology. Belmont, CA: WADSWORTH Cengage Learning.
- Huff, J.D., & Nietfeld, J.L. (2009). Using strategy instruction and confidence judgments to improve metacognitive monitoring. *Metacognition Learning*, 4, 161 – 176.
- Hyland, F., & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing*, 10, 185 – 212.
- Ibabe, I., & Juaregizar, J. (2010). Online self-assessment with feedback and metacognitive knowledge. *Higher Education*, 59, 243 – 258.
- Kaplan, A. (2008). Clarifying metacognition, self-regulation, and self-regulated learning: What's the purpose? *Educational Psychology Review*, 20, 477 – 484.
- Kardash, C. M., & Amlund, J. T. (1991). Self-reported learning strategies and learning from expository text. *Contemporary Educational Psychology*, 16(2), 117-138.
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational measurement: Issues and practice*, 30(4), 28-37.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254.

- Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. *The American Journal of Occupational Therapy*, 45, 214–222.
- Lajoie, S. P. (2008). Metacognition, self-regulation, and self-regulated learning: A rose by any other name? *Educational Psychology Review*, 20(4), 469-475.
- Leki, I., Cumming, A., & Silva, T. (2008). *A synthesis of research on second language writing in English*. Routledge: New York
- Lewins, A., & Silver, C. (2007). Using software in qualitative research: A step-by-step guide. Thousand Oaks, CA: SAGE Publications
- Li, J., & Barnard, R. (2011). Academic tutors' beliefs about and practices of giving feedback on students' written assignments: A New Zealand case study. *Assessing Writing*, 16(2), 137-148.
- McCormick, C.B. (2003). Metacognition and learning. In W.M. Reynolds, G.E. Miller, I.B. Weiner (Eds) Handbook of Psychology, Vol. 7, *Educational Psychology* (pp. 79 - 102). New York: John Wiley and Sons. Netlibrary
- McInerny, G.M., Walker, R.A., & Liam, A.D. (2011). Sociocultural theories of learning and motivation: Looking back, Looking forward. In G.M. McInerny, R.A. Walker, & A.D. Liam (Eds.) *Sociocultural theories of learning and motivation* (pp. 3 - 10). Charlotte, NC: Information Age Publishing.
- McManus, S. (2008). *Attributes of effective formative assessment*. Council of Chief State School Officers.
- McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 22(4), 34-43.

- McMillan, J.H. (2010). The practical implications of educational aims and contexts for formative assessment. In H. Andrade & G. Cizek (Eds.), *Handbook of formative assessment* (pp.41 – 58). New York, NY: Routledge
- McMillan, J.H., Venable, J., & Varier, D. (2013). Studies of the effect of formative assessment on student achievement: So much more is needed. *Practical Assessment Research and Evaluation*, 18(2), 2 available at <http://www.pareonline.net/pdf/v18n2.pdf>
- Metcalfe, J., & Finn, B. (2013). Metacognition and control of study choice in children. *Metacognition and Learning*, 8(1), 19-46.
- Miller, T. M., & Geraci, L. (2011). Training metacognition in the classroom: The influence of incentives and feedback on exam predictions. *Metacognition and Learning*, 6(3), 303-314.
- Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *The Journal of Experimental Education*, 74(1), 7 - 28.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218.
- Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *Tesol Quarterly*, 36(4), 573-595.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science*, 18(2), 119-144.
- Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1(2), 117-175.

- Piaget, J. (1970). *Science of education and the psychology of the child* (D. Coltman, Trans.). New York, NY: Orion Press.
- Pintrich, P. R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology*, 25(1), 92-104.
- Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice*, 41(4), 219 - 225.
- Poehner, M. E. (2011). Dynamic assessment: Fairness through the prism of mediation. *Assessment in Education: Principles, Policy & Practice*, 18(2), 99-112.
- Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education*, 34(1), 1 - 20.
- Pryor, J., & Crossouard, B. (2010). Challenging formative assessment: disciplinary spaces and identities. *Assessment & Evaluation in Higher Education*, 35(3), 265-276.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28(1), 4-13.
- Ruiz-Primo, M.A. & Li, M. (2013). Examining formative feedback in the classroom context: New research perspectives. In J.H. McMillan (Ed.), *SAGE Handbook of research on classroom assessment* (pp. 215 - 232). Thousand Oaks, CA: SAGE
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science*, 18(2), 119-144.
- Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in education*, 5(1), 77-84.

- Sardareh, S.A., & Saad, M.R.M. (2012). A sociocultural perspective on Assessment for Learning: The case of a Malaysian primary school ESL context. *Procedia- Social and Behavioral Sciences*, 66,343- 353.
- Schellings, G., & Wolters, B.V. (2011). Measuring strategy use with self-report instruments: theoretical and empirical considerations. *Metacognition Learning*, 6, 83 – 90.
- Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education*, 2(3), 114-121.
- Schoen, L.T. (2011). Conceptual and methodological issues in sociocultural research and theory development in Education. In G.M. McInerny, R.A. Walker, & A.D. Liam (Eds.) *Sociocultural theories of learning and motivation* (pp. 11 - 42). Charlotte, NC: Information Age Publishing.
- Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, 4(1), 33-45.
- Schraw, G. (2010). Measuring metacognitive judgments. In D.J. Hacker, J. Dunlosky, & A.C. Graesser (Eds.) *Handbook of metacognition in education* (pp. 415 - 429). New York: NY: Routledge.
- Schraw, G., & Denison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19(4), 460 - 475.
- Schunk, D.H. (2008). Metacognition, self-regulation, and self-regulated learning: Research recommendations. *Educational Psychology Review*, 20, 463-467.

- Scrimsher, S. & Tudge, J. (2003). The teaching/learning relationship in the first years of school: Some revolutionary implications of Vygotsky's theory. *Early Education & Development*, 14(3), 293 – 312.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational researcher*, 4-14.
- Spada, N., & Frohlich, M. (1995). COLT. Communicative Orientation of Language Teaching Observation Scheme: Coding conventions and applications. Sydney, Australia: National Centre for English Language Teaching and Research.
- Sperling, R. A., Howard, B. C., Staley, R., & DuBois, N. (2004). Metacognition and self-regulated learning constructs. *Educational Research and Evaluation*, 10(2), 117-139.
- Soslau, E. (2012). Opportunities to develop adaptive teaching expertise during supervisory conferences. *Teaching and Teacher Education*, 28(5), 768-779.
- Supovitz, J. A. (2012). Getting at student understanding: The key to teachers' use of test data. *Teachers College Record*, 114(11), 1-29.
- Tabachnick, B. G., & Fidell, L. S. (2007). Using multivariate statistics. Boston, MA: Allyn & Bacon
- Timmers, C. F., Braber-Van Den Broek, J., & Van Den Berg, S. M. (2013). Motivational beliefs, student effort, and feedback behaviour in computer-based formative assessment. *Computers & Education*, 60(1), 25-31.
- Tobias, S., & Everson, H. T. (2009). The importance of knowing what you know: A knowledge monitoring framework for studying metacognition in education. In D.J. Hacker, J. Dunlosky, & A.C. Graesser (Eds.) *Handbook of metacognition in education* (pp. 107-127). New York, NY: Routledge.

- Torrance, H. (1993). Formative assessment: Some theoretical problems and empirical questions. *Cambridge Journal of Education*, 23(3), 333.
- Tudge, J., & Schrimsher, S. (2010). Lev S. Vygotsky on education: A cultural-historical, interpersonal, and individual approach to development. In B. Zimmerman & D.H. Schunk (Eds.). *Educational Psychology: A Century of Contributions* (pp 207 – 228). New York: Routeledge
- Veenman, M.V.J. (2011). Alternative assessment of strategy use with self-report instruments: A discussion. *Metacognition Learning*, 6, 205 – 211.
- Veenman, M.V.J., Hout-wolters, B. , & Afflerbach, P. (2006). Metacognition and learning: conceptual and methodological considerations. *Metacognition and Learning*, 1, 3 – 14.
- Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA. Harvard University Press.
- Weir, C. (1998). Using Embedded Questions to Jump-Start Metacognition in Middle School Remedial Readers. *Journal of Adolescent & Adult Literacy*, 41(6), 458-67.
- White, B., & Frederiksen, J. (2005). A theoretical framework and approach for fostering metacognitive development. *Educational Psychologist*, 40(4), 211-223.
- Wiliam, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. Andrade & G. Cizek (Eds.), *Handbook of formative assessment* (pp.18 – 40). New York, NY: Routledge.
- Wiliam, D. (2013). Feedback and instructional correctives. In J.H. McMillan (Ed.), *SAGE Handbook of research on classroom assessment* (pp. 197 - 214). Thousand Oaks, CA: SAGE

- William, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education*, 11(1), 49-65.
- William, D., & Leahy, S. (2007). A theoretical foundation for formative assessment. In J.H. McMillan (Ed.), *Formative classroom assessment: Theory and practice* (pp. 29 - 42). Thousand Oaks, CA: SAGE
- Willis, J. (2011). Affiliation, autonomy and Assessment for Learning. *Assessment in Education: Principles, Policy & Practice*, 18(4), 399-415
- Winne & Hadwin, 2000
- Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P.R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (p. 532 - 566). San Diego, CA: Academic Press.
- Yin, R. (2014). Case Study Research: Design and Methods. Thousand Oaks, CA: Sage Publications.
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, 45(4), 477-501.
- Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, 81(3), 329.
- Zimmerman, B. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25(1), 3 – 17.

Appendices

Appendix A

ESL Teacher Assessment Practices Interview Protocol

PARTICIPANT ID:

DATE:

INTERVIEW NUMBER

SECTION 1: Background Questionnaire

Gender: Male Female

Years of Teaching Experience:

Years of ESL teaching experience:

Highest Education Level:

Courses taught in ESL contexts:

Current class size:

Briefly describe a typical day in a level 3 writing class.

Formalized Training in Assessment (coursework or professional development workshops related to developing, analyzing, and using assessments):

None Minimal Moderate Extensive

Please rate your skills in developing assessments: Very weak Weak Strong Very strong

Please rate your skills in making adjustments to instruction on an ongoing basis: Very weak Weak
Strong Very strong

SECTION 2: Grading and Assessment Practices

1. Could you describe your assessment and grading practices in your ESL classes?

Note: Clarify that assessment practices refer to type (teacher made, curriculum based, textbook based, etc.), frequency, and nature of assessments used to evaluate learning; may include graded and non-graded tasks

2. Could you describe/what is your plan for assessments in the level 3 written communication class?

Note: Ask about daily, weekly, monthly assessments as relevant

3. To what extent are your assessment and grading practices influenced by expectations and curriculum of the English Language Program? Not at all Some Extent Moderate Extent Major Extent

--Could you elaborate on your choice?

Note: Clarify that by grading you mean assigning performance criteria and evaluating learning by use of numeric scores/letter grades

SECTION 3: Teacher beliefs about the nature of learners PART 1

4. What proportion of students do you typically expect to receive an 'A' in a semester? This semester?
--What do you think are the characteristics of an 'A' student in terms of motivational/engagement characteristics as well skills?

Note: The participant may be asked to describe a B, C, D student as well. Participant may also be asked to elaborate on motivational characteristics

5. What do you think are the characteristics of a student who is struggling in terms of motivational/engagement characteristics as well skills?
--What proportion of students do you expect to receive an 'F' in a typical semester? This semester?

SECTION 4: Ongoing/Informal Formative Assessment Practices

LEARNING GOALS

6. How are learning goals or objectives communicated to students?
-- How do you think students, if they do, understand learning goals?
-- What do you think is the role of learning goals in successful learning? In this case, learning level 3 writing?

ELICITING EVIDENCE OF STUDENT UNDERSTANDING

7. In your daily instructional practices, how do you know that students have understood/are understanding the content/skills that you are teaching?

-- Are there explicit and implicit clues that may indicate difficulty/ease for students during a lesson? If yes, could you give a few examples?

INSTRUCTIONAL ADJUSTMENTS (NARROWING THE GAP BETWEEN CURRENT UNDERSTANDING AND GOALS)

8. In your daily instructional practices, how often do you adjust instruction based on evidence that students know/don't know the content/skills that you are teaching?

Never/Rarely Sometimes Often Always

Note: Clarify instructional adjustments are making revisions to planned lesson, changing content/skills, increase/decrease time allotted for a topic, etc.

-- Could you elaborate on your choice?

9. In your daily instructional practices, how often do you provide feedback to students on their understanding?

Never/Rarely Sometimes Often Always

-- Please elaborate on your choice

-- Could you describe your feedback practices in your daily interactions with students in this class? *Follow up: written/oral; class/group/individual; spontaneous/planned; immediate/delayed*

SECTION 5: Use of assessment performance data to inform instruction

Section Introduction: This section includes questions related to how you use formal assessments to inform your instructional practices. Formal assessments are those that are given periodically and resemble formal tests (essay or multiple choice) rather than informal question or activities done in the classroom.

10. How do you use student performance information from formal assessments (including quizzes, tests, chapter tests, homework exercises) to know whether students have understood what you have taught?

11. How often do you adjust instruction based on evidence from formal assessment that students know/don't know the content/skills that you taught?

Never/Rarely Sometimes Often Always

Could you elaborate on your choice?

SKIP Questions 12 – 14 if participant says he/she does not use formal assessments for formative purposes

12. How often do you provide feedback on student performance in formal assessments?

Never/Rarely Sometimes Often Always

13. Could you describe your feedback practices based on performance in formal assessments with students in this class? *Follow up: written/oral; class/group/individual; spontaneous/planned; immediate/delayed*

14. What are your perceptions of the utility of classroom interactions versus formal assessments in providing information on how well/ how much students have learned the content/skills?

Note: Clarify that the intent is to compare and contrast on the type, depth of information on student understanding, if necessary: how is the information different? Which one is more useful?

SECTION 6: Teacher beliefs about the nature of learners PART 2

15. What do you think are particular student characteristics that influence how students use feedback to improve their content knowledge or skills in writing?

16. Could you describe how students typically use feedback given by you?

17. Could you describe your general perceptions of the metacognitive skills of the students you teach?

Note: Clarify or provide description of metacognitive skills to include student awareness of their knowledge, skills of content and strategies and their ability to regulate strategies to learn.

- Is it important to teach these skills explicitly? What is your role in improving students' metacognitive skills?

SECTION 7: Conclusion

Would you like to talk about anything related that we have not covered in the interview?

Conclusion: Thank you for taking the time to participate in this interview. I may contact you to check if the transcriptions and interpretations from this interview are consistent with what you intended to communicate. Please do not hesitate to contact if you have any questions or concerns regarding this interview.

Appendix B

ESL Teacher Assessment Practices Interview Protocol – II

PARTICIPANT ID:

DATE:

INTERVIEW NUMBER:

SECTION 1: Background

On a typical day, what instructional activities did you employ in the level 3 written communication OR describe a typical day in a level 3 writing class.

Average preparation time for each class period:

Please rate your skills in developing assessments: Very weak Weak Strong Very strong

Please rate your skills in making adjustments to instruction on an ongoing basis: Very weak Weak Strong Very strong

SECTION 2: Grading and Assessment Practices

Summarize teacher response on the first interview.

1. Based on our last interview, have your grading and assessment practices changed since the last interview. If yes, in what ways?

2. To what extent were your grading practices for this class influenced by expectations and curriculum of the English Language Program for the level 3 writing class?

Not at all Some Extent Moderate Extent Major Extent

--Could you elaborate on your choice?

SECTION 3: Beliefs about nature of learners

3. Based on the first interview, have your perceptions of the characteristics of a student who gets an 'A' or 'F' grade in this class and F in this class changed? If yes, could you describe the changes for each?

Note: Summarize previous response to the two items

4. Based on the student survey data, how would you describe this class of students as learners?

5. Based on the eight judgments of learning questionnaires, how would you describe this class of students as learners?

Note: Clarify that participant may talk about motivation, engagement, ability, and other learner characteristics

SECTION 4: Ongoing/Informal Formative Assessment Practices

LEARNING GOALS

6. Were there any changes in how learning expectations or goals for level 3 writing communicated to students? (*Summarize previous interview response to the question*)

ELICITING EVIDENCE OF STUDENT UNDERSTANDING

7. In your daily instructional practices for this class, how did you know that students understood/were understanding the content/skills that you were teaching?

INSTRUCTIONAL ADJUSTMENTS (NARROWING THE GAP BETWEEN CURRENT UNDERSTANDING AND GOALS)

8. In your daily instructional practices, how often did you adjust instruction based on evidence that students knew/don't know the content/skills that you were teaching?

Never/Rarely Sometimes Often Always

Note: Clarify instructional adjustments are making revisions to planned lesson, changing content/skills, increase/decrease time allotted for a topic, etc.

8b. Could you elaborate on your choice?

9. Could you describe your feedback practices in your daily interactions with students in this class?

Follow up: written/oral; class/group/individual; spontaneous/planned; immediate/delayed

10. Based on the student survey data provided to you at the beginning of the semester, did you make any modifications to your daily instructional practices during the semester? If yes, in what ways? If no, why?

Note: follow up may be related to whether they reflected on the data, if they discussed it with students, etc.

SECTION 5: Use of assessment performance data to inform instruction

Section Introduction: This section includes questions related to how you use formal assessments to inform your instructional practices. Formal assessments are those that are given periodically and resemble formal tests (essay or multiple choice) rather than informal question or activities done in the classroom.

11. How did you use student performance information from formal assessments (including quizzes, tests, chapter tests, homework exercises) to know whether students have understood what you taught?

12. How often did you adjust instruction based on evidence from formal assessment that students know/don't know the content/skills that you taught?

Never/Rarely Sometimes Often Always

13b. Could you elaborate on your choice?

13. Did you see a pattern in the frequency/nature of these adjustments or use in relation to the time of the semester? (e.g., Beginning/Middle/End)

14. Could you describe your feedback practices based on performance in formal assessments with students in this class? *Follow up: written/oral; class/group/individual; spontaneous/planned; immediate/delayed*

15. What are your perceptions of the utility of classroom interactions versus formal assessments in providing information on how well/ how much students have learned the content/skills?

Note: Clarify that the intent is to compare and contrast on the type, depth of information on student understanding, if necessary: how is the information different? Which one is more useful?

16. Based on the student data provided to you, did you make any changes to the nature, frequency or type of feedback you provided to students during the semester?

SECTION 6: Teacher beliefs about the nature of learners PART 2

17. What do you think influences whether and to what extent instructional adjustments and feedback help students improve their content knowledge or skills in writing?

18. How do you think your feedback practices and instructional adjustments helped students improve their knowledge/skills?

19. Could you describe your general perceptions of the motivational beliefs of the students you teach?
-- How does motivation affect your students' behavior in class and performance on assessments?

Note: Clarify or provide description of motivation to include students' level of interest, attitudes, effort in learning

20. Could you describe your general perceptions of the metacognitive skills of the students you teach?
22b. How do metacognitive skills affect your students' behavior in class and performance on assessments?

Note: Clarify or provide description of metacognitive skills to include student awareness of their knowledge, skills of content and strategies and their ability to regulate strategies to learn.

21. Based on the student data, how consistent or different was the data in relation to your previous experiences and beliefs about level 3 written communication students?

Note: Follow up questions may include probes on specific aspects of the questionnaire like confidence levels, difficulty perceptions, etc.

SECTION 7: Conclusion

Would you like to talk about anything related to this topic that we have not covered in the interview?

Conclusion: Thank you for taking the time to participate in this interview. I may contact you to check if the transcriptions and interpretations from this interview are consistent with what you intended to communicate. Please do not hesitate to contact if you have any questions or concerns regarding this interview.

Appendix C

ESL Student Metacognitive Beliefs Survey

Instructions

The purpose of the survey is to understand your thoughts and feelings about learning to write in English so that your teacher can know more about how you feel about learning English and adjust teaching to help you learn better.

There is no right or wrong answer to any question in the survey. Please respond as honestly as possible. You may skip any question that makes you feel uncomfortable. Only the researcher of this study can see your answers to the survey.

Your teacher will get a summary of the survey results. This means that the researcher will provide information of how the students answered the survey questions. The teacher will not be able to identify how you or your classmates responded to the survey individually.

Taking the survey is voluntary, and your participation or non-participation will not affect your grade in any way.

It will take approximately 15 - 30 minutes to complete the survey. If you have any questions or concerns, you may contact the researcher at the phone number or email address below:

Researcher's Name: Divya Varier

Email: varierd@vcu.edu

Phone: 757-819-8103

1. Age:

- Below 18
- 18 – 25
- 26 – 35
- 36 – 45
- Above 45

2. Gender:

- Male
- Female

3. Student status:

- Full time ELP
- Full time undergraduate
- Full time graduate
- Other

4. Country of Origin:

5. How often to you converse with your friends or family in English?

Never Rarely Sometimes Often Always

6. As a English language learner, how difficult is each of the following English language skills listed below?

SKILLS	Not at all difficult	Somewhat Difficult	Difficult	Very Difficult
Reading				
Writing				
Listening				
Speaking				

10. Please respond to the following based on the extent to which you agree or disagree with each statement.

	Strongly Disagree	Disagree	Neither Disagree Nor Agree	Agree	Strongly Agree
I have a good understanding of English grammar					
I have a good understanding of the components of an essay					
When I read a paragraph in English, I can understand the main points					
I can summarize what I read in English					
When writing in English, I pay attention to the words I am using					
When writing in English, I pay attention to grammar					
When writing in English, I pay attention to the requirements of the assignment or test					
I am good at writing in English					

18. Please respond to the following based on the extent to which you agree or disagree with each statement.

	Strongly Disagree	Disagree	Neither Disagree Nor Agree	Agree	Strongly Agree
Reading my classmates' writing helps me improve my own writing					
Providing feedback on my classmates' writing helps me improve my own writing					
I am good at writing in my native language					
Writing well is important for other subjects					
Written communication (emails, letters, etc.) is easier than writing for coursework					

23. Please respond to the following based on the extent to which you agree or disagree with each statement.

	Strongly Disagree	Disagree	Neither Disagree Nor Agree	Agree	Strongly Agree
When writing in English, I try to use what I learned in class					
I try to apply ideas from course readings in class activities like discussion or exercises					

If you would like to share anything about your experience and beliefs on learning English please use the text box below:

Thank you for taking the survey

Appendix D

ESL Formative Assessment Observation Guide

Teacher ID:
Date & Time:
Observation Number:

Duration:

Setting

Number of students in class:

Investigator notes on student behavioral engagement:

SECTION 2A: SUMMARY OF ACTIVITIES

ACTIVITY	COUNT	APPROXIMATE DURATION OF EACH (MINUTES)	DISCOURSE*	NOTES
INDIVIDUAL				
GROUP				
CLASS				
WRITING				
READING				
LECTURE				
OTHER				

Discourse: NONE/ MINIMAL/MODERATE/EXTENSIVE FOR THE DURATION OF ONE ACTIVITY.

SECTION 2: NOTES

TO BE FILLED BASED ON FIELD NOTES

SECTION 3: FORMATIVE ASSESSMENT INTERACTIONS

Origin	CHECK
Primarily teacher-initiated	
Primarily student-initiated	
combination	
Process attributes	
Criteria for success: Learning goals/expectations are communicated to students (McManus, 2008).	
Feedback: <i>specificity</i>	
Individual (given to individual student)	
Group (given to a group of students)	
Check student comprehension (right/ wrong)	
Elaborate on student understanding.	
Offers suggestions on how student can improve.	
Feedback linked to learning goals/expectations.	
Feedback: <i>type</i>	
Corrective (provides correct answer)	
Metacognitive - elicit student thought (Panova & Lyster , 2002)	
Feedback: <i>timing</i>	
Students given time to reflect on feedback before making changes/revisions.	
Uptake: <i>Student response to feedback</i> (Panova & Lyster, 2002)	
Student/s responded verbally (appropriate)	
Student/s responded verbally (incorrect/inappropriate)	
Student/s did not respond	

Adapted from McMillan, Venable, & Varier, 2013

Appendix E

Metacognitive Judgments of Learning Questionnaires

Date:

Student Name:

Instructions

The purpose of this questionnaire is to understand your preparation and how you feel about the material on Unit 6. There is no right or wrong answer, and your responses **will not** affect your grade in any way. Please answer the questions as honestly as possible.

Before beginning the test, please answer the following questions on a scale of 1 – 10 by circling the number most appropriate for you.

1. How well did you prepare for today's test?

(did not prepare at all) 1 ----2----3----4----5----6--7----8----9----10 (prepared very well)

2. How well do you know what was taught in Unit 6?

(not well at all) 1 ----2----3----4----5----6----7----8----9----10 (very well)

3. How difficult is the unit test given to you today?

(Not at all difficult) 1-----2-----3-----4-----5-----6-----7-----8-----9-----10 (very difficult)

4. How confident are you that you will do well in this unit test?

(Not at all confident) 1 ----2----3----4----5----6----7----8----9----10 (very confident)

OPTIONAL:

Please use the space below if you would like to share any thoughts or feelings about today's test.

Appendix F

JOL-Performance Summary Report

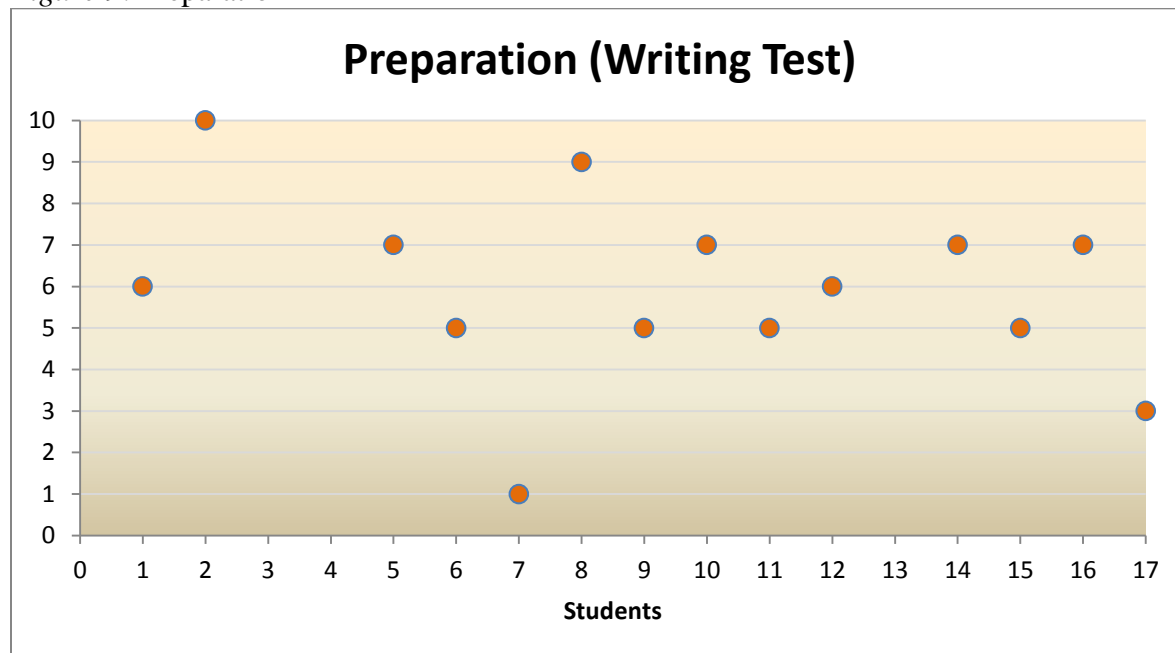
Summary of Student Responses to the Judgments of Learning Questionnaire – Unit 6 Writing & Skills Test

The figures below summarize the results from the administration of the judgments of learning (JOL) questionnaire for students in **WC 3 section 001**. The numbers on the X axis refer to each student of the class based on the class roster – so, student 1 in figures 9 - 20 is the first student on the class roster. The chart may be interpreted based on patterns for the class or the individual student. Some charts allow interpretation of student ratings and their performance at the individual and class level.

Writing Test

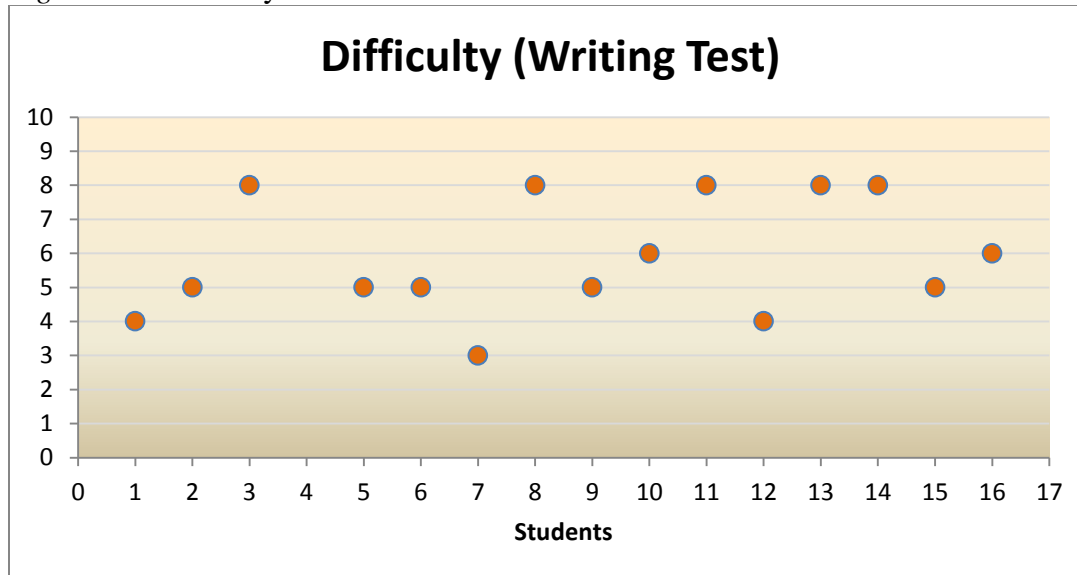
For the JOL on preparation (figure 9), there was considerable variation in student responses ranging from 1 (did not prepare at all) to 10 (prepared very well). The average rating for the group was 5.64. However, a majority of students rated their preparation at or above 5.

Figure 9: Preparation



For the JOL on difficulty of the writing test, thirteen students responded to the item. A majority of students did not perceive the test to be difficult (average 5.8) – 9 selected a range of 4 -6; five students selected a difficulty level of 8/10, and one student selected 3/10.

Figure 10: Difficulty



For the JOL on confidence (figure 11 below), the class average was 6.3. Students in general reported high level of confidence - 8/17 students selected a rating between 4 and 6, and 6/17 students selected a rating between 7 and 10. Two students selected a rating of 3.

Figure 11: Confidence

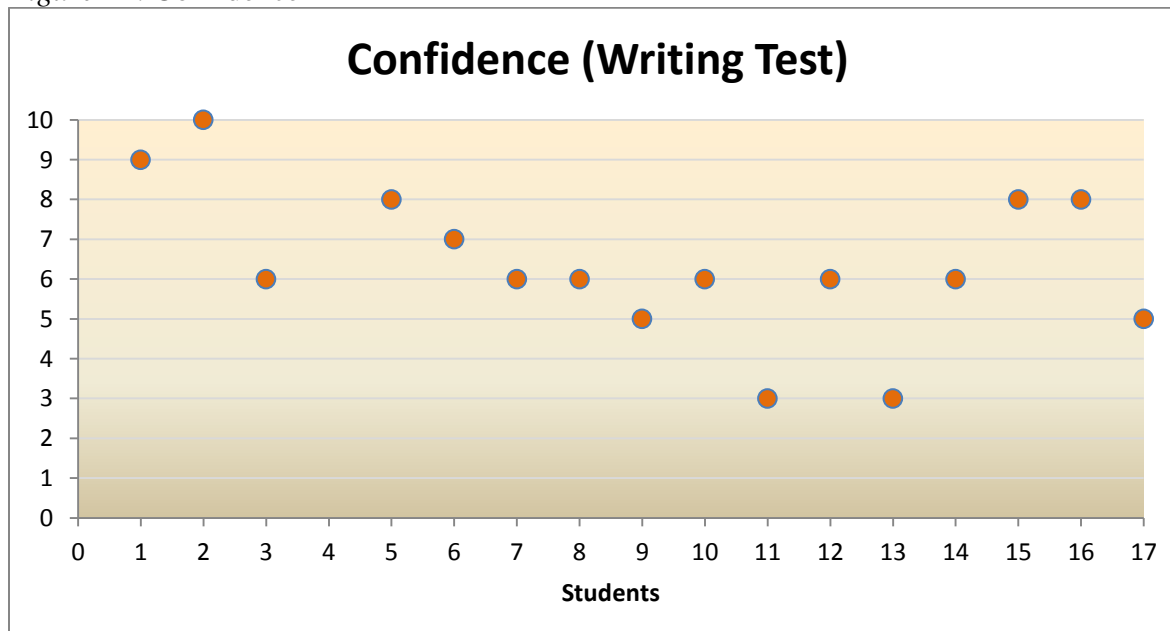
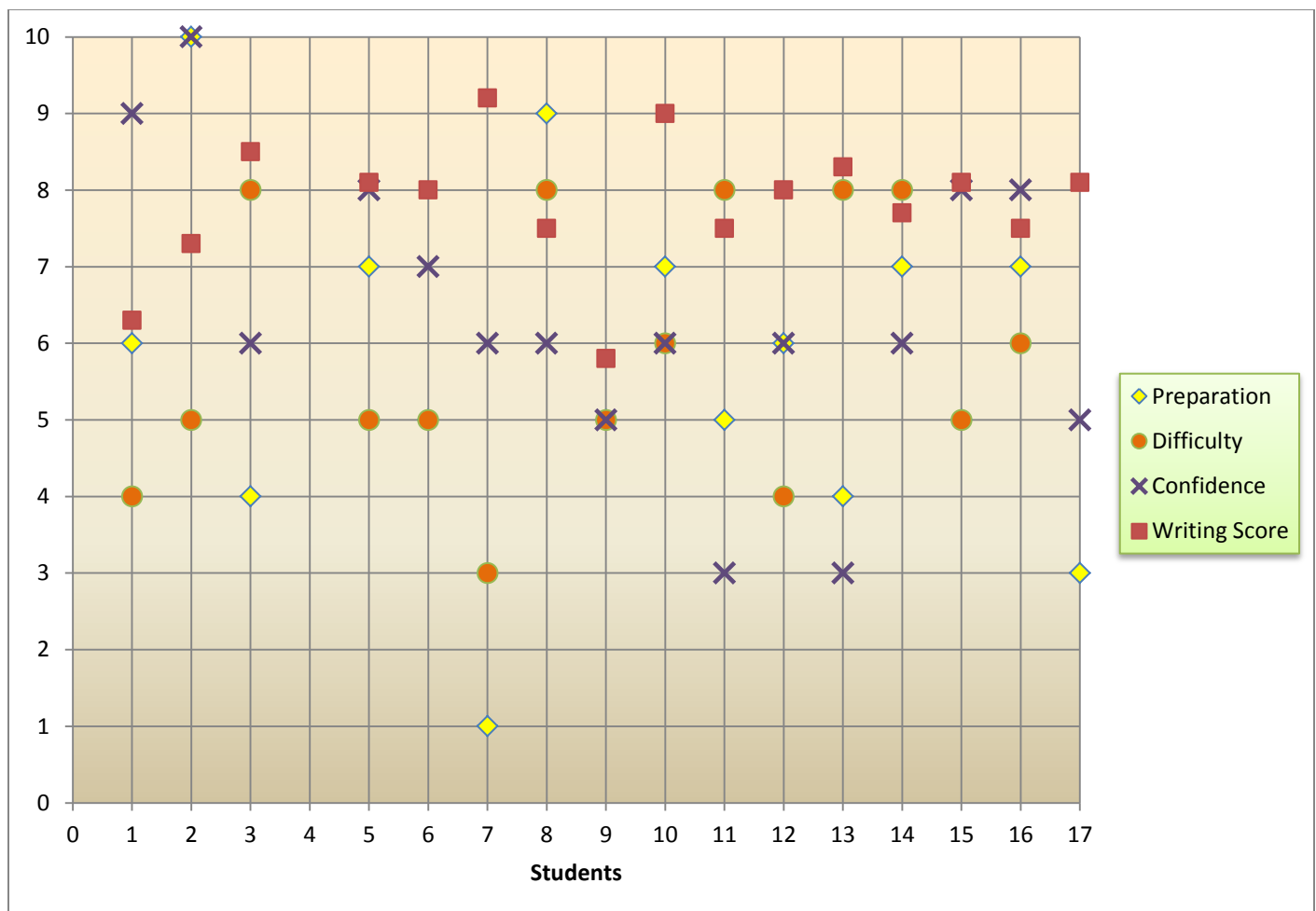


Figure 12 provides an illustration of the three JOL in relation to students' performance on the writing test. A significant correlation was not found between student performance and JOL. However, the two students (student number 7 and 10) who received an A in the writing test showed under confidence with a rating of 6; student 7 reported not preparing for test, and a low difficulty level (3/10); student 10 reported a similar rating for confidence and preparation (6-7), and a difficulty rating of 6. Students who received a B on the writing test showed similar ratings of under confidence with the exception of two students whose confidence ratings matched their writing performance. For students who received a C grade or below, students scoring mid - upper C grade were similar to A and B students, however two students with a low C and below, reported they were very confident about doing well in the writing test. There was no clear pattern in the relationship between JOL on difficulty and performance- for example, students who did not perform well did not perceive the test to be difficult.

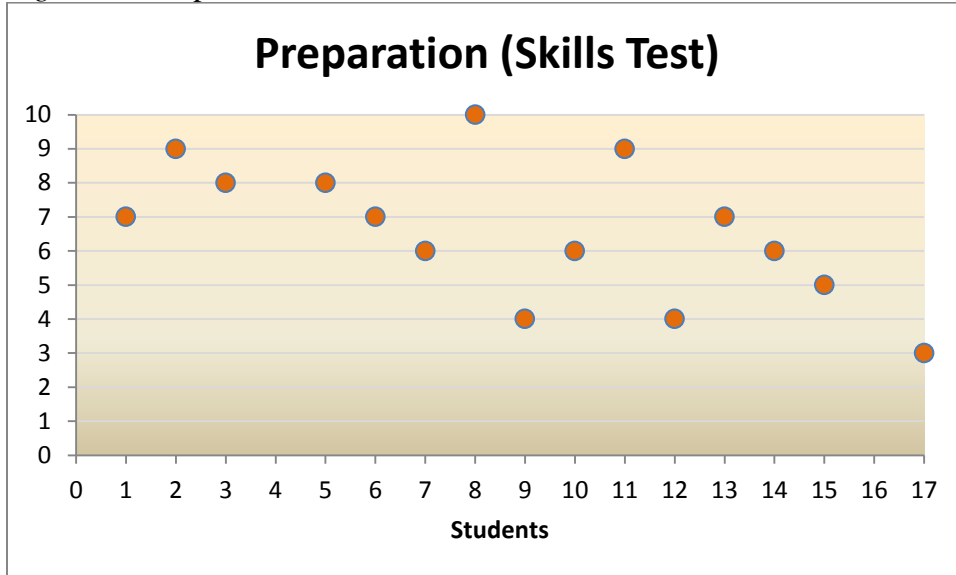
Figure12: JOL & Performance: Writing test



Skills Test

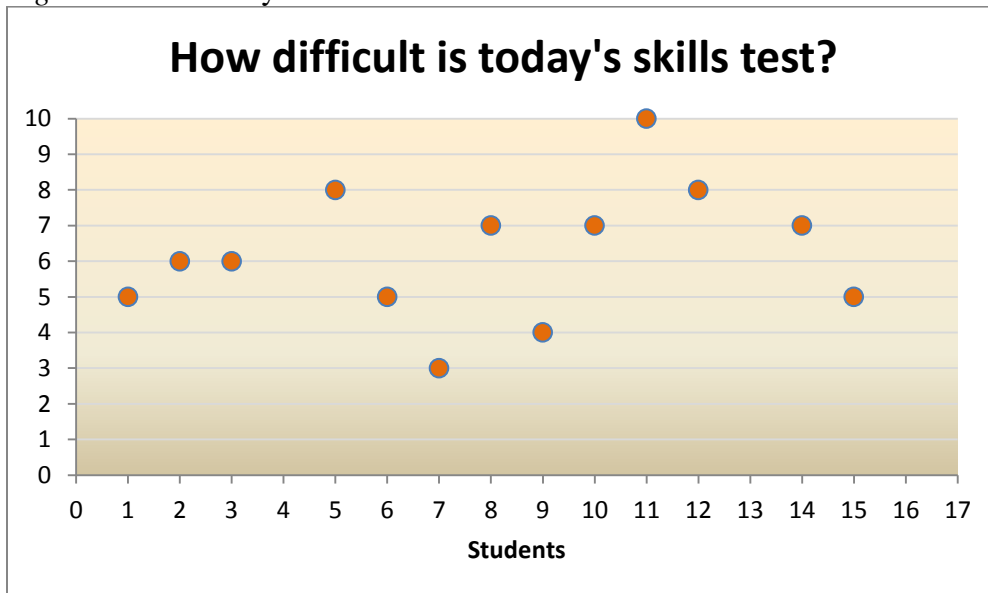
For JOL on preparation, the average rating for the class was about one point higher than the preparation for writing at 6.6/ 10. Only one student provided a rating of 3, indicative of low preparation, and a majority of students rated above 6 on this JOL.

Figure 13: Preparation for Skills Test



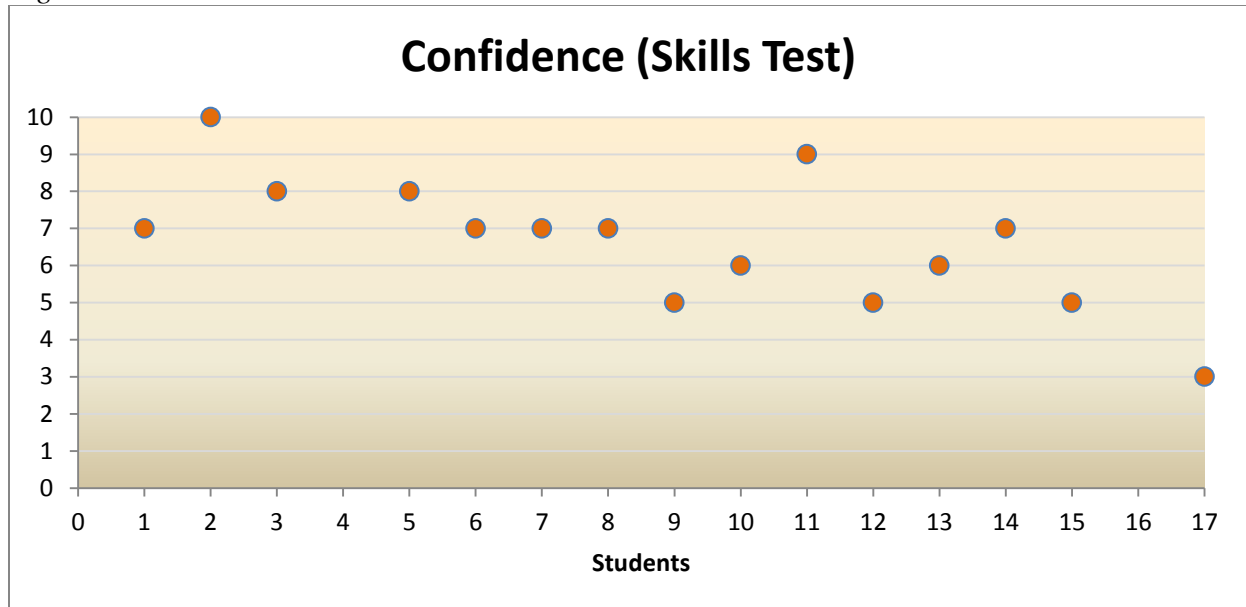
For test difficulty, students perceived this test to be slightly more difficult than the writing test (average = 6.2). and about 6 students rated difficulty at or above 7.

Figure 14: Difficulty for Skills test



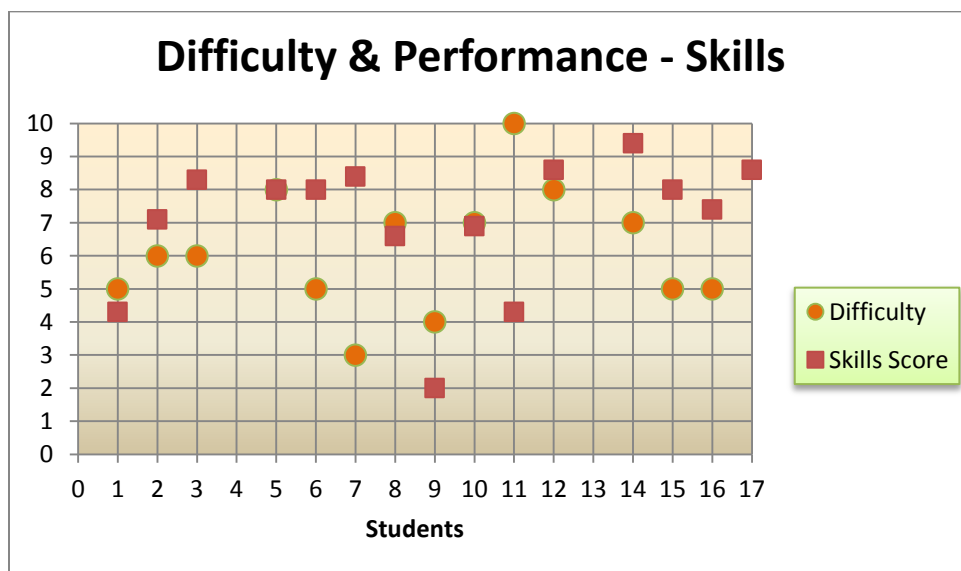
On average, the class rated a confidence level of 6.7, with a majority of students reporting a confidence of 7 or above.

Figure 15: Confidence for skills test



With the exception of one student (student 11) who rated the test to be very difficult, most students perceived the test to be fairly difficult regardless of their performance. For example, Student 12 rated the test difficulty at 8, and received a B grade, whereas Student 1 who received a grade below C rated test difficulty at 5.

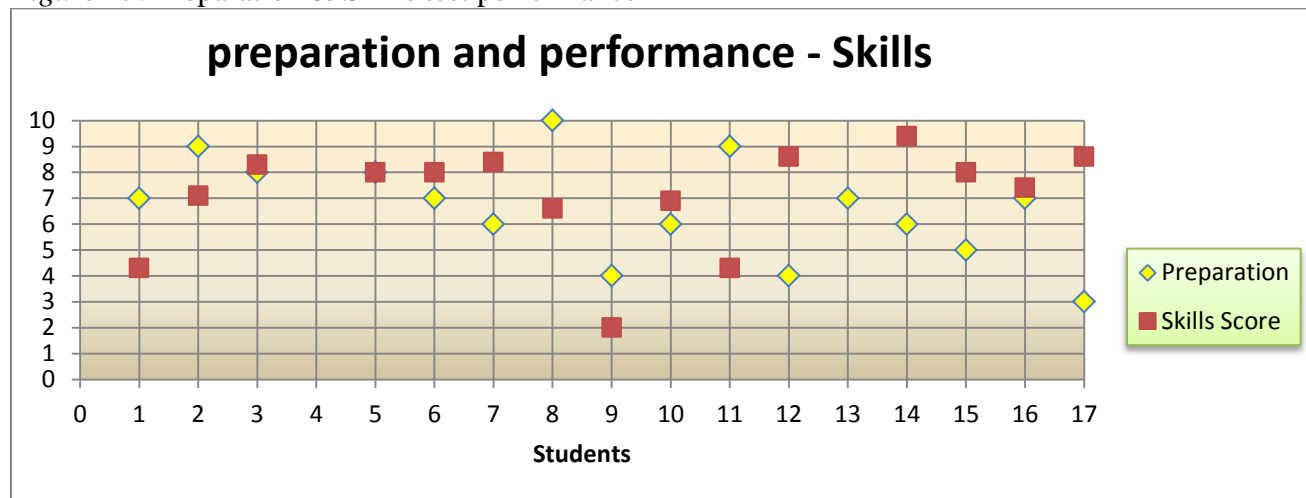
Figure 16: JOL & Performance – Difficulty & Skills



JOL & Performance – Preparation

Although a majority of students reported they prepared fairly well or very well (6- 10) range, their performance showed more variation with some students who reported they prepared well scoring below C, and some students who reported a lower rating in preparation receiving an A or B grade.

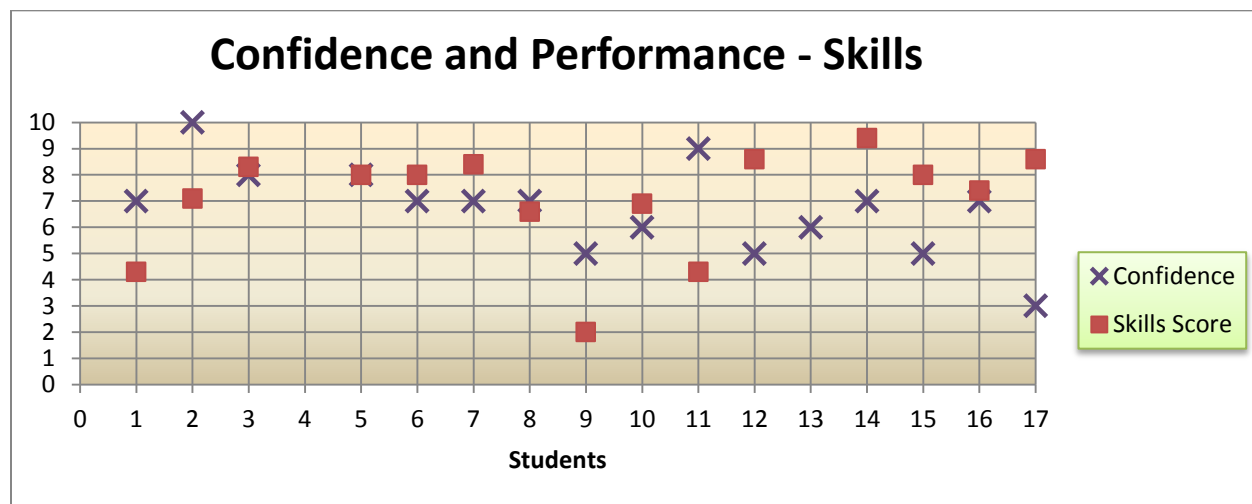
Figure 17: Preparation & Skills test performance



Confidence & Skills test scores

Almost all students reported feeling fairly or very confident about how well they would perform in the skills test with the exception of one student who reported low confidence (3); in relation to their performance, students who received an A or a B showed some under confidence whereas students who received lower grades tended to report they were either fairly confident or very confident.

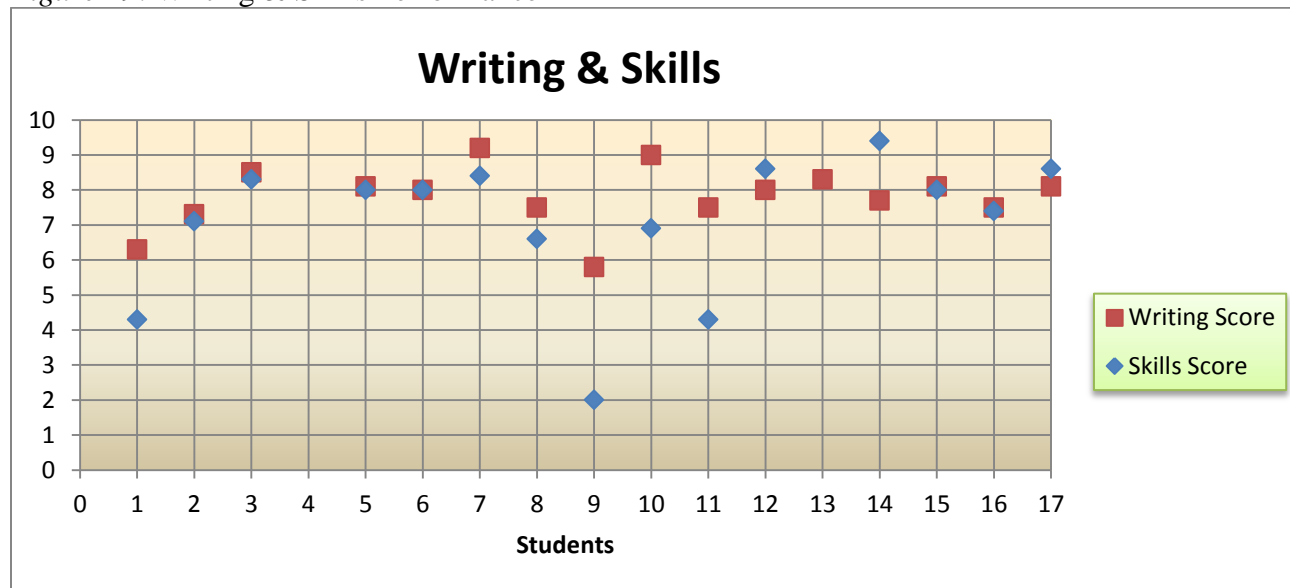
Figure 18: JOL & Performance – Confidence



Performance in writing and skills test

As expected, there is a strong positive association between students' writing and skills score; In general, students who performed well in the writing also did well in the skills test.

Figure 19: Writing & Skills Performance



The knowledge JOL asked students to report on the extent to which they felt they knew the material taught in Unit 6. Students responded to this item on the writing test day as well as the skills test day. The following chart provides an illustration of student JOL on this item the day before the skills test, and on the day of the skills test, in relation to their performance on the skills tests. This analysis was based on the reasoning that the unit 6 material related closely to the skills test. For each student, we can look at their perception of their knowledge before skills test day, and on skills test day. In general, students reported they knew the material fairly well or very well (above 7), but students varied in their ratings in relation to their performance- some students' ratings were close to their performance, and others under estimated or over-estimated their knowledge of the material. See figure 20 below.

Appendix G

List of Codes

ATLAS.ti output: Initial Code List

Code Family "Setting"

Teacher Planning Time
ESL courses taught
Educational qualification
Context: classroom physical setting
Obs classroom physical characteristics
Proportion getting A grade
Proportion getting F grade
Training in Assessment
Perceived training in assessment
Years of teaching experience
Perceived skills: assessment development

Code Family "Situation or Context"

Obs classroom physical characteristics
Context: classroom physical setting
Context: Collaboration with teachers
Grading Practices
Context: Influence of Program Expectations
Context: NonTeaching Role
Obs Class Characteristics
Context: Flexibility related to assessments
Assessment Practices
Obs vocabulary
Class characteristics
Obs References to Culture or Language
Engaging with Student Culture
Context: Role of student culture
Obs managing classroom behavior
Obs Classroom Interaction
Obs Lateness
Obs Attendance

Code Family "Activity Codes"

Obs writing portfolio: reflection
Peer Group Activity

Homework Review
Follow-up:Typical Day
Obs Writing Portfolio: Drafts
Obs individual activity
Writing Portfolio: Drafts
Obs Peer or group Activity
Obs Review of Assignments
Obs class activity
Typical Day Beginning of Semester
Writing Portfolio: Reflection
Practice exercises
Obs individual conference
Obs practice exercises
Obs homework review
Individual conference

Code Family "Student"

Proportion getting F grade
Proportion getting A grade
Obs Lateness
Obs Classroom Interaction
Obs Attendance
Characteristics of an A student
Student characteristics
Student characteristics: Educational/Academic Experiences
Characteristics of an F student
Obs Class Participation

Code Family "Teacher Beliefs"

Years of teaching experience
Perceived skills: assessment development
Training in Assessment
Perceived training in assessment
Attitudes: Assessment and Grading practices
Frustration
Perception: Self as teacher
Perception: JOL
Student improvement
Reflection on Instruction
Obs Frustration

Code Family "Strategies"

Obs teaching test taking strategies
Obs instructional adjustment
Use of Rubrics
Obs Use of Rubrics
Obs managing classroom behavior
Obs Instructional strategy
Informal or Formal: Use?
Feedback Strategies
Instructional Strategies

Code Family "Formative Assessment Theory"

Obs individual conference
Practice exercises
Individual conference
Obs practice exercises
Obs homework review
Obs Sharing Day's agenda
Obs Feedback Type Class level learning
Informal FA: Frequency of feedback
Obs Feedback type metacognitive
Peer Assessment: Use
Obs Specificity of feedback_Group
Obs Feedback_linked to learning goals
Obs teacher initiated FA
Instructional Adjustment
Obs Feedback Type Corrective
Obs Feedback Type Positive
Obs Peer Assessment Activity
Obs sharing grading criteria
Formal Assessment as Evidence of Learning
Sharing expectations/goals/objectives
Obs Response to Feedback_Verbal appropriate
Obs Response to Feedback_No Response
Formal FA: Nature of Feedback
Obs Informal FA: Any questions?
Obs Specificity of feedback_elaborate on student understanding
Obs specificity of feedback_suggestions on improving
Informal FA: check student understanding
Obs Specificity of feedback_check student understanding
Obs informal FA: group feedback
Informal FA: Specificity of Feedback
Obs student use of feedback
Obs Student Perception of Writing/Grade
Obs Teacher Reflection on Student Learning
Obs student initiated FA

Obs Response to Feedback_Verbal inappropriate
Obs Acknowledge Difficulty Ease
Student use of feedback
Formal FA: Specificity of feedback
Obs Feedback_time to revise
Obs Specificity of feedback_individual
Informal FA: Nature of Feedback
Obs Sharing learning expectations
Use of Rubrics
Obs instructional adjustment
Obs teaching test taking strategies
Obs Use of Rubrics

ATLAS.ti Output: Final Code List

Code Family "Activity Codes"

Follow-up: Typical Day
Homework Review
Obs homework review
Obs Peer or group Activity
Obs practice exercises
Peer Group Activity
Typical Day Beginning of Semester

Code Family "Situation or Context"

Assessment Practices
Context: classroom physical setting
Context: Collaboration with teachers
Context: Flexibility related to assessments
Context: Influence of Program Expectations
Obs classroom physical characteristics

Code Family "Teacher Characteristics"

Attitudes: Assessment and Grading practices
Context: NonTeaching Role
Educational qualification
ESL courses taught
Frustration
Obs Frustration
Obs managing classroom behavior
Perceived skills: assessment development
Perceived training in assessment
Perception: JOL
Perception: Self as teacher
Reflection on Instruction
Teacher Planning Time
Training in Assessment
Years of teaching experience

Code Family "Student Characteristics"

Characteristics of an A student
Characteristics of an F student
Class characteristics

Context: Role of student culture
Engaging with Student Culture
Obs Attendance
Obs Class Characteristics
Obs Class Participation
Obs Classroom Interaction
Obs Lateness
Obs References to Culture or Language
Proportion getting A grade
Proportion getting F grade
Student characteristics
Student characteristics: Educational/Academic Experiences

Code Family "Formative Assessment Theory"

Feedback Strategies
Formal Assessment as Evidence of Learning
Formal FA: Nature of Feedback
Formal FA: Specificity of feedback
Informal FA: check student understanding
Informal FA: Frequency of feedback
Informal FA: Nature of Feedback
Informal FA: Specificity of Feedback
Informal or Formal: Use?
Instructional Adjustment
Instructional Strategies
Obs Acknowledge Difficulty Ease
Obs Feedback Type Class level learning
Obs Feedback Type Corrective
Obs Feedback type metacognitive
Obs Feedback Type Positive
Obs Feedback_linked to learning goals
Obs Feedback_time to revise
Obs Informal FA: Any questions?
Obs informal FA: group feedback
Obs instructional adjustment
Obs Instructional strategy
Obs Peer Assessment Activity
Obs Response to Feedback_Verbal inappropriate
Obs Response to Feedback_No Response
Obs Response to Feedback_Verbal appropriate
Obs Sharing Day's agenda
Obs sharing grading criteria
Obs Sharing learning expectations
Obs Specificity of feedback_check student understanding
Obs Specificity of feedback_elaborate on student understanding

Obs Specificity of feedback_Group
Obs Specificity of feedback_individual
Obs specificity of feedback_suggestions on improving
Obs student initiated FA
Obs student use of feedback
Obs teacher initiated FA
Obs Teacher Reflection on Student Learning
Obs teaching test taking strategies
Obs Use of Rubrics
Peer Assessment: Use
Sharing expectations/goals/objectives
Student use of feedback
Use of Rubrics

VITA

Divya Varier was born on July 24, 1983, in Bangalore, India. She is a citizen of India. She completed her primary and secondary education in India, and received a Bachelor of Arts degree in Psychology, Economics, and English Literature from Mount Carmel College, Bangalore University in 2004. Her work experience includes print media advertising and counseling students pursuing advanced study in the United States. After moving to the United States, she received a Bachelor of Science degree majoring in Psychology, from Old Dominion University, Norfolk, VA, in 2010.

As a doctoral student, Divya served as a graduate assistant at the VCU School of Education. Her research interests are in the area of educational assessment, program evaluation, and international students' well-being and success. Divya has presented her work at local, national, and international educational research conferences organized by the VCU School of Education, VCU Graduate Student Association, American Educational Research Association (AERA), American Evaluation Association (AEA), and the International Association of Educational Assessment (IAEA).